

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 14, 2004, 09:09:48 ; Search time 66 Seconds
(without alignments)
4096.939 Million cell updates/sec

Title: US-09-917-376-1
Perfect score: 5135
Sequence: 1 MDRSENIRLTMRSRRLVSLT.....RASFGSVNPATPTADTVLQX 957

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues
Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5134	100.0	957	ABP73015	Abp73015 Amino aci
2	5134	100.0	957	ADD22821	Add22821 Acidother
3	4036	78.6	740	ABP73016	Abp73016 Amino aci
4	4036	78.6	740	ADD22926	Add22926 Acidother
5	4036	78.6	740	ADD22923	Add22923 Acidother
6	2174	42.3	940	ABP99489	Abp99489 Amino aci
7	1680	32.7	726	ADD22927	Add22927 Aspergill
8	1625.5	31.7	838	ADD42055	Add42055 Trichoder
9	848	16.5	555	ADD24919	Add24919 Xanthomon
10	823	16.0	1228	ABP73029	Abp73029 Amino aci
11	702	13.7	762	ABP73018	Abp73018 Amino aci
12	468	9.1	88	ABP73017	Abp73017 Amino aci
13	468	9.1	88	ADD22825	Add22825 Acidother
14	462	9.0	88	ADD22825	Add22825 Acidother
15	462	9.0	89	ADD22924	Add22924 Acidother
16	462	9.0	154	ABP73020	Abp73020 Amino aci
17	459	8.9	150	ABP73025	Abp73025 Amino aci
18	429	8.4	153	ABP71658	Abp71658 A. cellul
19	429	8.4	1121	ABP71656	Abp71656 A. cellul
20	334	6.5	1751	AAV13493	AAV13493 Truncated
21	334	6.5	1751	AAE16324	AAE16324 Active ce
22	323	6.3	782	AAE15625	AAE15625 Cellulase
23	301.5	5.9	1749	AAE72314	AAE72314 Glutamic
24	288	5.6	96	ABP78539	Abp78539 Gum arabi
25	276	5.4	1426	AAV13492	AAV13492 Truncated

26	276	5.4	1426	5	AAE16323	AAE16323 Active ce
27	274	5.3	406	4	ABG27250	Abg27250 Novel hum
28	262	5.1	2468	6	ABU38411	Abu38411 Protein e
29	262	5.1	2468	6	ABP59933	Abp59933 Microbial
30	261.5	5.1	531	2	AAW15238	AAW15238 Scaffold1
31	261.5	5.1	1853	2	AAW43108	AAW43108 C. thermo
32	257.5	5.0	732	5	ABP93616	Abp93616 Herbicida
33	253	4.9	2435	6	ABU19529	Abu19529 Protein e
34	244	4.8	1664	2	AAW43106	AAW43106 C. thermo
35	243.5	4.7	1049	6	ABU21976	Abu21976 Protein e
36	243	4.7	1259	5	ABP857561	Abp857561 Human MC2
37	238	4.6	200	5	ABP78541	Abp78541 TIR-Pro-T
38	237	4.6	103	5	ABP78540	Abp78540 Ala-Pro-A
39	234	4.6	2137	5	ABP39618	Abp39618 Staphyloc
40	232.5	4.5	1016	6	ABU41567	Abu41567 Protein e
41	229.5	4.5	950	4	ABP71513	Abp71513 Drosophil
42	227.5	4.4	455	4	AAU43436	AAU43436 Propionib
43	227.5	4.4	455	6	ABM40865	Abm40865 Propionib
44	227.5	4.4	2271	6	ABU16000	Abu16000 Protein e
45	226.5	4.4	1831	6	ABU43109	Abu43109 Protein e

ALIGNMENTS

RESULT 1
ABP73015 standard; protein; 957 AA.
ID ABP73015
AC ABP73015;
DT 03-JUN-2003 (first entry)
XX Amino acid sequence of the avicelase AviiiI.
XX Avicelase; AviiiI; glycoside hydrolase; enzyme; cellulase; biofuel;
XX detergent; pulp processing; paper processing; feed processing; textile;
XX cellulose.
XX Acidothermus cellulolyticus.
OS
FH Key Location/Qualifiers
FT Misc-difference 957
FT /note="unspecified residue encoded by N"
XX
XX W02003012090-A2.
XX
XX 13-FEB-2003.
XX
XX 28-JUL-2001; 2001MO-US023818.
XX
XX 28-JUL-2001; 2001MO-US023818.
XX
XX (MIDE) MIDWEST RES INST.
XX
XX Ding S, Achey WS, Vinzant TB, Himmel ME,
XX WPI; 2003-248177/24.
XX N-FSDB; ABZ77632.
XX
XX New thermostable AviiiI peptide from Acidothermus cellulolyticus, useful
XX for degradation of cellulose or in generating anti-AviiiI antibodies for
XX purifying recombinant AviiiI polypeptides from genetically engineered
XX host cells.
XX
XX Claim 2; Page 20; 4pp; English.
XX
XX The present sequence represents a thermostable avicelase polypeptide,
XX designated AviiiI. AviiiI is a member of the glycoside hydrolase family
XX of enzymes, and is a cellulase. AviiiI is useful in the conversion of
XX biomass to biofuels and biofuel additives. It may be useful in the
XX production of detergents, pulp and paper processing, food and feed
XX processing and in textile processes. The thermostable AviiiI peptide is


```
QY 121 PINTNKVAAGVMTNSWDPNDAILRSSDQATWQITPLPKLGANNMGRGGERLAVD 180
DB 121 PINTNKVAAGVMTNSWDPNDAILRSSDQATWQITPLPKLGANNMGRGGERLAVD 180
QY 181 PNNNNILYFGAPSGKGLMRSTDSGATWQMTNFPVGVITYIANPTDTTQGSQSDIQGVVWA 240
DB 181 PNNNNILYFGAPSGKGLMRSTDSGATWQMTNFPVGVITYIANPTDTTQGSQSDIQGVVWA 240
QY 241 FDKSSSSILGQASKITFVGADPNPNPVFWSRDGATWQAVPAGPTGFI PHKGVFDPVNHVL 300
DB 241 FDKSSSSILGQASKITFVGADPNPNPVFWSRDGATWQAVPAGPTGFI PHKGVFDPVNHVL 300
QY 301 YIATSNTEGPDYSGSDVWKFVSITSGTWTRISPVSTDTANDYFGYSGLTIDROHNTIM 360
DB 301 YIATSNTEGPDYSGSDVWKFVSITSGTWTRISPVSTDTANDYFGYSGLTIDROHNTIM 360
QY 361 VATQISWMPDITIIFRSTDTGATWTRIMWTSYPNRSLRYVLDISAEPMULTFQVQNPVPV 420
DB 361 VATQISWMPDITIIFRSTDTGATWTRIMWTSYPNRSLRYVLDISAEPMULTFQVQNPVPV 420
QY 421 SPKLGMDDEAMAIIDPNSDRMLYGTATLYATNDLTKMDSGQIHIAVWKLGEETAVND 480
DB 421 SPKLGMDDEAMAIIDPNSDRMLYGTATLYATNDLTKMDSGQIHIAVWKLGEETAVND 480
QY 481 LISPPSGAPLISALGDLGFTHADYTAVPSTIFTSPTVTTGTSVDYAEINPISIIYRAGSF 540
DB 481 LISPPSGAPLISALGDLGFTHADYTAVPSTIFTSPTVTTGTSVDYAEINPISIIYRAGSF 540
QY 541 DPSSQPNDRHVAFTSDGKKNMFQSGEPGVTTGTVAASADGSRFVMAAGDPQGVVYAV 600
DB 541 DPSSQPNDRHVAFTSDGKKNMFQSGEPGVTTGTVAASADGSRFVMAAGDPQGVVYAV 600
QY 601 GFGSMASAGVAPNAQIRSDRVNPKTEYALSNGFYVSTGCGTFFQVPAAGLPSSGANG 660
DB 601 GFGSMASAGVAPNAQIRSDRVNPKTEYALSNGFYVSTGCGTFFQVPAAGLPSSGANG 660
QY 661 VMFPAVNGKEGDLMLAASSGLYHSTNGSSWSAITGVSSAAVVGKSGAPSSYPAVFV 720
DB 661 VMFPAVNGKEGDLMLAASSGLYHSTNGSSWSAITGVSSAAVVGKSGAPSSYPAVFV 720
QY 721 GTTGGVTGAYRSDCGTTWVLINDOHQYGNMGQAITGDHANLRRVYIGTNGRGIVYGD 780
DB 721 GTTGGVTGAYRSDCGTTWVLINDOHQYGNMGQAITGDHANLRRVYIGTNGRGIVYGD 780
QY 781 GGAPESSPSPSVSASPSLSPPSPSSPSPSPSSPSPSPSSPSPSPSPSPSPSPSPSP 840
DB 781 GGAPESSPSPSVSASPSLSPPSPSSPSPSPSSPSPSPSSPSPSPSPSPSPSPSPSP 840
QY 841 SASPSSPSPSPSSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 900
DB 841 SASPSSPSPSPSSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 900
QY 901 SSVDLSVTVRYWTRTGGSSSTLYVNCMAAIGCNIRASFGSVAPATPTADTYIQ 956
DB 901 SSVDLSVTVRYWTRTGGSSSTLYVNCMAAIGCNIRASFGSVAPATPTADTYIQ 956

RESULT 3
ABP73016 ID ABP73016 standard; peptide; 740 AA.
AC ABP73016;
XX 03-JUN-2003 (first entry)
DE Amino acid sequence of the avicelase AvilIII catalytic domain.
XX
XX
XX Avicelase; AvilIII; glycoside hydrolase; enzyme; cellulase; biofuel;
XX detergent; pulp processing; paper processing; feed processing; textile;
XX cellulose.
OS Acidothermus cellulolyticus.
```

```
XX
XX WO2003012090-A2.
XX
XX 13-FEB-2003.
XX
XX 28-JUL-2001; 2001MO-US023818.
XX
XX 28-JUL-2001; 2001MO-US023818.
XX
XX (MIDE ) MIDWEST RES INST.
XX
XX Ding S, Adney WS, Vinzant TB, Himmel ME.
XX
XX WPI; 2003-248177/24.
XX
XX New thermostable AvilIII peptide from Acidothermus cellulolyticus, useful
XX for degradation of cellulose or in generating anti-AvilIII antibodies for
XX purifying recombinant AvilIII polypeptides from genetically engineered
XX host cells.
XX
XX Claim 6; Page 8; 44pp; English.
XX
XX The present sequence is derived from a thermostable avicelase, designated
XX AvilIII. AvilIII is a member of the glycoside hydrolase family of enzymes,
XX and is a cellulase. AvilIII is useful in the conversion of biomass to
XX biofuels and biofuel additives. It may be useful in the production of
XX detergents, pulp and paper processing, food and feed processing and in
XX textile processes. The thermostable AvilIII peptide is useful in the
XX degradation of cellulose, and in generating specific anti-AvilIII
XX antibodies that are useful in purifying recombinant AvilIII polypeptides
XX from genetically engineered host cells, in detecting AvilIII polypeptide
XX expression, as well as a reagent tool for characterizing the molecular
XX actions of the polypeptide. The AvilIII polynucleotide is useful as a
XX source of probes or primers in various diagnostic assays
XX
XX Sequence 740 AA:
SQ
Query Match 78.6%; Score 4036; DB 6; Length 740;
Best Local Similarity 100.0%; Pred. No. 3.2e-205;
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
47 ATTQPTWNSVAIGGGFVDGIVFNEGAGILVYRTDIGMYMDANGRMIBLDMVGM 106
DB 1 ATTQPTWNSVAIGGGFVDGIVFNEGAGILVYRTDIGMYMDANGRMIBLDMVGM 60
QY 107 NNWGVNVSIADPINTNKVAAGVMTNSWDPNDAILRSSDQATWQITPLPKLG 166
DB 61 NNWGVNVSIADPINTNKVAAGVMTNSWDPNDAILRSSDQATWQITPLPKLG 120
QY 167 NMPGRMGERLAVDPNNDNILYFGAPSGKGLMRSTDSGATWQMTNFPVGVITYIANPTDT 226
DB 121 NMPGRMGERLAVDPNNDNILYFGAPSGKGLMRSTDSGATWQMTNFPVGVITYIANPTDT 180
QY 227 TGYQSDIQGVVWVAFPKSSSSILGQASKITFVGADPNPNPVFWSRDGATWQAVPAGPTGF 286
DB 181 TGYQSDIQGVVWVAFPKSSSSILGQASKITFVGADPNPNPVFWSRDGATWQAVPAGPTGF 240
QY 287 IPHKGVFDPVNHVLIATSNTEGPDYSGSDVWKFVSITSGTWTRISPVSTDTANDYFGY 346
DB 241 IPHKGVFDPVNHVLIATSNTEGPDYSGSDVWKFVSITSGTWTRISPVSTDTANDYFGY 300
QY 347 SGLTIDROHNTIMVATQISWMPDITIIFRSTDTGATWTRIMWTSYPNRSLRYVLDISAE 406
DB 301 SGLTIDROHNTIMVATQISWMPDITIIFRSTDTGATWTRIMWTSYPNRSLRYVLDISAE 360
QY 407 PMULTFQVQNPVPVPSKLGMDDEAMAIIDPNSDRMLYGTATLYATNDLTKMDSGQIH 466
DB 361 PMULTFQVQNPVPVPSKLGMDDEAMAIIDPNSDRMLYGTATLYATNDLTKMDSGQIH 420
QY 467 APMVWKLGEETAVNDLISPPSGAPLISALGDLGFTHADYTAVPSTIFTSPTVTTGTSVDY 526
DB 421 APMVWKLGEETAVNDLISPPSGAPLISALGDLGFTHADYTAVPSTIFTSPTVTTGTSVDY 480
```

QY 527 AELNPSIIIVRAGSPDSSQPNDRHVAFTDGGKMFQGSSEPGVTTGGTVAASADGSRFV 586
 DB 481 AELNPSIIIVRAGSPDSSQPNDRHVAFTDGGKMFQGSSEPGVTTGGTVAASADGSRFV 540
 QY 587 WAPGDPGQPVVYAVGFGNSMAASQGVPAANAQIRSDRNPVKTFFALSNCTFFRSTDDGVTF 646
 DB 541 WAPGDPGQPVVYAVGFGNSMAASQGVPAANAQIRSDRNPVKTFFALSNCTFFRSTDDGVTF 600
 QY 647 QPVAAGLPSSGAVGVMEFHAVPKGKEDLMLAASGLYHSTNGGSSWSAITGVSSAVNVFG 706
 DB 601 QPVAAGLPSSGAVGVMEFHAVPKGKEDLMLAASGLYHSTNGGSSWSAITGVSSAVNVFG 660
 QY 707 KSAFGSSYPVAVFVGTIGVTVGAYRSDDCGTTWVLINDQHQYGNMGQAITGDHANLRV 766
 DB 661 KSAFGSSYPVAVFVGTIGVTVGAYRSDDCGTTWVLINDQHQYGNMGQAITGDHANLRV 720
 QY 767 YIGTNGRGIVYGDIGGAPSG 786
 DB 721 YIGTNGRGIVYGDIGGAPSG 740

RESULT 4

ADD22926
 ID ADD22926 standard; protein; 740 AA.

AC ADD22926;

DT 15-JUN-2004 (first entry)

XX Acidothermus cellulolyticus avicelase AvIII catalytic domain.

KM enzyme; AvIII; cellulose reduction; agricultural biomass;

KM municipal solid waste; glycoside hydrolase; avicelase.

OS Acidothermus cellulolyticus.

XX US2003108988-A1.

PD 12-JUN-2003.

PF 18-OCT-2002; 2002US-00155400.

PR 28-JUL-2001; 2001US-00917376.

XX (DING/) DING S.

PA (ADNE/) ADNEY W S.

PA (VINZ/) VINZANT T B.

PA (HIMM/) HIMMEL M E.

PI Ding S, Adney WS, Vinzant TB, Himmel ME;

XX WPI; 2003-810853/76.

XX New isolated thermal tolerant avicelase polynucleotide useful for

XX detection of a polynucleotide encoding AvIII and for reducing cellulose

XX in a starting material, e.g. municipal solid waste.

XX Example 2; SEQ ID NO 6; 29pp; English.

XX The invention relates to an isolated polynucleotide molecule encoding a

XX thermostable AvIII polypeptide. The polynucleotide is useful for

XX detection of a polynucleotide encoding AvIII. The polynucleotide is

XX useful for reducing cellulose in a starting material which involves

XX administering to the starting material, e.g. agricultural biomass or

XX municipal solid waste, a polypeptide molecule of the polynucleotide. The

XX method further comprises administering a second polypeptide molecule

XX chosen from the glycoside hydrolase family of proteins. The present

XX sequence represents the amino acid sequence of Avicelase. The present

XX cellulolyticus avicelase AvIII catalytic domain.

XX Sequence 740 AA;

XX Query Match 78.6%; Score 4036; DB 7; Length 740;

Best Local Similarity 100.0%; Pred. No. 3.2e-205; Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 ATTOPYTMSNVAIGGGFVDGIVFNEGAPGIIYVRFDIGMYRMDANGRMIPLDVMG 106
 DB 1 ATTOPYTMSNVAIGGGFVDGIVFNEGAPGIIYVRFDIGMYRMDANGRMIPLDVMG 60
 QY 107 NNMGVNVSIADPINTNKVAAVGMVYNSMDPNDGAILRSSDQATWQITPLPKLG 166
 DB 61 NNMGVNVSIADPINTNKVAAVGMVYNSMDPNDGAILRSSDQATWQITPLPKLG 120
 QY 167 NMPGGMGERLLAVDPNDNIIYFGAPSSGKLRSTDSGATWSQMTNPPVGVYTIANPTDT 226
 DB 121 NMPGGMGERLLAVDPNDNIIYFGAPSSGKLRSTDSGATWSQMTNPPVGVYTIANPTDT 180
 QY 227 TGYOSDIOGVVWVAFPKSSSSIGQASKTIFVGVADPNNVFPMRSDGATWQAVGAPG 286
 DB 181 TGYOSDIOGVVWVAFPKSSSSIGQASKTIFVGVADPNNVFPMRSDGATWQAVGAPG 240
 QY 287 IPHKGVPDPVNHVLYATSNVTGGPYDSSGQVWKFVSTSGTWRISPVSTDPANDYFGY 346
 DB 241 IPHKGVPDPVNHVLYATSNVTGGPYDSSGQVWKFVSTSGTWRISPVSTDPANDYFGY 300
 QY 347 SGLTIDRQHPNTIMVATQISWMPDTIIFRSTDGATWIRIMWTSYIPNRLRYVDISAE 406
 DB 301 SGLTIDRQHPNTIMVATQISWMPDTIIFRSTDGATWIRIMWTSYIPNRLRYVDISAE 360
 QY 407 PMLTFGVQNPVPSPKLGMDDEAMAIDPNSDRMLYGTATLYATNDLTKMDSGQIHT 466
 DB 361 PMLTFGVQNPVPSPKLGMDDEAMAIDPNSDRMLYGTATLYATNDLTKMDSGQIHT 420
 QY 467 APWVKGLEETAANDLISPPSGAPLISALGDLGFTHADVTAVPSTIFTSVFETGTSVDY 526
 DB 421 APWVKGLEETAANDLISPPSGAPLISALGDLGFTHADVTAVPSTIFTSVFETGTSVDY 480
 QY 527 AELNPSIIIVRAGSPDSSQPNDRHVAFTDGGKMFQGSSEPGVTTGGTVAASADGSRFV 586
 DB 481 AELNPSIIIVRAGSPDSSQPNDRHVAFTDGGKMFQGSSEPGVTTGGTVAASADGSRFV 540
 QY 587 WAPGDPGQPVVYAVGFGNSMAASQGVPAANAQIRSDRNPVKTFFALSNCTFFRSTDDGVTF 646
 DB 541 WAPGDPGQPVVYAVGFGNSMAASQGVPAANAQIRSDRNPVKTFFALSNCTFFRSTDDGVTF 600
 QY 647 QPVAAGLPSSGAVGVMEFHAVPKGKEDLMLAASGLYHSTNGGSSWSAITGVSSAVNVFG 706
 DB 601 QPVAAGLPSSGAVGVMEFHAVPKGKEDLMLAASGLYHSTNGGSSWSAITGVSSAVNVFG 660
 QY 707 KSAFGSSYPVAVFVGTIGVTVGAYRSDDCGTTWVLINDQHQYGNMGQAITGDHANLRV 766
 DB 661 KSAFGSSYPVAVFVGTIGVTVGAYRSDDCGTTWVLINDQHQYGNMGQAITGDHANLRV 720
 QY 767 YIGTNGRGIVYGDIGGAPSG 786
 DB 721 YIGTNGRGIVYGDIGGAPSG 740

RESULT 5

ADD22923
 ID ADD22923 standard; protein; 740 AA.

AC ADD22923;

DT 15-JUN-2004 (first entry)

XX Acidothermus cellulolyticus avicelase AvIII catalytic domain.

XX enzyme; AvIII; cellulose reduction; agricultural biomass;

XX municipal solid waste; glycoside hydrolase; avicelase.

XX Acidothermus cellulolyticus.

XX US2003108988-A1.

XX	12-JUN-2003.
XX	
PF	18-OCT-2002; 2002US-00155400.
XX	
FR	28-JUL-2001; 2001US-00917376.
XX	
PA	(DING/) DING S.
PA	(ADNEY/) ADNEY W S.
PA	(VINZ/) VINZANT T B.
PA	(HIMM/) HIMMEL M E.
XX	
FI	Ding S, Adney WS, Vinzant TB, Himmel ME;
XX	
DR	WPI; 2003-810853/76.
XX	
PT	New isolated thermal tolerant avicelase polynucleotide useful for
PT	detection of a polynucleotide encoding AvIII and for reducing cellulose
PT	in a starting material, e.g. municipal solid waste.
XX	
XX	Claim 16; SEQ ID NO 3; 29pp; English.
XX	
CC	The invention relates to an isolated polynucleotide molecule encoding a
CC	thermostable AvIII polypeptide. The polynucleotide is useful for
CC	detection of a polynucleotide encoding AvIII. The polynucleotide is
CC	useful for reducing cellulose in a starting material which involves
CC	administering to the starting material, e.g. agricultural biomass or
CC	municipal solid waste, a polypeptide molecule of the polynucleotide. The
CC	method further comprises administering a second polypeptide molecule
CC	chosen from the glycoside hydrolase family of proteins. The present
CC	sequence represents the amino acid sequence of Acidothermus
CC	cellulolyticus avicelase AvIII catalytic domain.
XX	
QX	Sequence 740 AA;

QY	587	WAPGPGQPVVYAAVFGSGMAASQCVPNAAQJRSRVRNPKTFYALNSNGFFYSTGGYTF	646
Db	541	WAPGPGQPVVYAAVFGSGMAASQCVPNAAQJRSRVRNPKTFYALNSNGFFYSTGGYTF	600
QY	647	QPVAAGLPSSGAVGVGMFHAVPEKEKEDLWLAASSGLYHSTNGSSSMWAIIGVSAVNVGFG	706
Db	601	QPVAAGLPSSGAVGVGMFHAVPEKEKEDLWLAASSGLYHSTNGSSSMWAIIGVSAVNVGFG	660
QY	707	KSAPSSSYPAVFVVGTTIGVYTGAYRSDDCGTTWVLINDQHOYGNMGOAITGDHANLRV	766
Db	661	KSAPSSSYPAVFVVGTTIGVYTGAYRSDDCGTTWVLINDQHOYGNMGOAITGDHANLRV	720
QY	767	YIGTNGRGIVYGDIGAPSG 786	
Db	721	YIGTNGRGIVYGDIGAPSG 740	
RESULT 6			
ID	ABB99489	standard; protein; 940 AA.	
AC	ABB99489;		
DT	03-MAR-2003	(first entry)	
DE	Amino acid sequence of a xyloglucanase enzyme.		
XX			
KW	Xyloglucanase; enzyme; family 74; glycosyl hydrolase; cellulosic fiber;		
XX			
KM	textile scouring; cellulose fiber processing; ratting.		
XX			
OS	Jonesia sp.		
FN	WO200277242-A2.		
PD	03-OCT-2002.		
XX			
PP	27-MAR-2002; 2002WO-DK000210.		
PR	27-MAR-2001; 2001DK-00000504.		
XX			
PA	(NOVO) NOVOZYMES AS.		
XX			
P1	Duffner F, Sjöholm C;		
DR	WPI; 2003-092855/08.		
DR	N-PSDB; ABV76941.		
XX			
PS	Claim 5; Page 72-75; 76pp; English.		
XX			
CC	The present sequence represents a xyloglucanase enzyme, belonging to		
CC	family 74 of glycosyl hydrolases. The enzyme is isolated from <i>Jonesia</i> sp.		
CC	DSM14140. The enzyme is useful in processes for machine treatment of		
CC	fabrics. It is also useful in the textile industry for improving the		
CC	properties of cellulosic fibers, yarn, woven or non-woven fabric, and in		
CC	a textile scouring process step. The xyloglucanase enzyme is also useful		
CC	in the cellulose fiber processing industry for ratting of fibers such as		
CC	hemp, jute, flax and linen. It is useful for preventing binding of		
CC	certain soils to the xyloglucan left on the cellulosic material		
QY	Sequence 940 AA;		

QY 70 FNEGAPGILYRTDYGMYRMDANGRMIPLDWVGANNNGVVSIAADPINTKWA 129
 DB 62 FNGKDPGLVYARTDIGAYRLNDSTGRWIPLDHIGMDWSHGILSLATDEVDTRVYL 121
 QY 130 AVGWYTSMDNDGAILSSDQATWQITPLPKLGGMMPGMGEBRLAVDNNNDILYF 189
 DB 122 AAGTYSDMDQNGAILRSADKGETMELPFRVGNNGMGEBRLAVDNNNDILYF 181
 QY 190 GAPGKGLMRSTDSGATWSQMTNFPDVGTYIANPDTTG-YQSDIQGVVVAFDKSSSL 248
 DB 182 GABGNGIMKSTDKYKTKTSFPAAGNYA--DAGATYGNQGVVWTFDPISAA 238
 QY 249 GQAKTIFVGVADPNPNPVFWSRDGATWQAVGAPGTGFLPKHGVDPVNVHLYIATSN 308
 DB 239 GQTTQITIVGVADKNNVYRSTDGATWQVPGGTGFLAQGVLDHKQGYIATSDTG 298
 QY 309 GPYGGSGDWKFSVTSQTRISFVSTDTANDYFGISGLTIIDKHPNTIWAIOISW 368
 DB 299 GPYDGSKDWRLDISGQWTRISPIST-SSNSAFVSGLAIDKKNPDTIIVVSVQVSW 357
 QY 369 PDTIIFRSTGGATWTRIMDWTYSNRLRYVLDISAEPLTFEGVQPNPVSPKLGMD 428
 DB 358 PDMYVRSIDKGTWSPIMELNGSOPRTQYHNDISGAPWDFGTAKGPRANPKLGMD 417
 QY 429 EAMADFNDRMLYGTGATLYATNDLTWKMDGQIHLAPWVGLSETAVNDLISPSGA 488
 DB 418 QSFIDPNDRFPFGTGTAGIYGNLNMMDKGVDTIVKQGLETAQDLAPNGMI 477
 QY 489 PLISALDGLGFTADVTAVPT-IFTSPTVTTGTSVDYAEINPSTIIVAGSPDPSSQN 547
 DB 478 GLYALADIGGFTHDISQVPRKYKKNPHDTVISIDFAESKPAVVAAGK--SISGRT 535
 QY 548 DRHAFSTGDKWQSGSEPGVTTGTVAAADSRRFMAACDQGPVYVAVGPNNSA 607
 DB 536 TSWGVSTDAGETWPKATPGVKGPGSITVSANASSIYAP--EGAAPRRSTNSGSSWS 593
 QY 608 ASQGVANAGVIRSDPNKFTFYALNGTFYRSTDDGVTFOPA-AGLPSSGAVGWFAV 666
 DB 594 TVSGLPFNAQVADRNVANTLYGFVDGKFYHSTNGGASTABAFGTFTSG--NARFRAV 651
 QY 667 PGKGDMLA-----ASSGLYHSTNGGSSMSAITGVSSAVNVGFGKAPGSSYPAVVG 721
 DB 652 PGRGHLMLAGVSGSTYGMWRSTGDKNMTKVSQVQEDAVGFKATSSSGYPIVFTA 711
 QY 722 TIGGVTAAYRSDCTWVLINDDOHYGNMGOALTGDHANLRVYITNGKGIYVGDIG 781
 DB 712 KIDVARGIFRSDDDEKTKRINDNHNQWMTGASITGPDVGRVYITNGRGIYGD-- 769
 QY 782 GAPGSPSPVSPASPLSP 840
 DB 770 -----STPPPPGQDSGASGNTGATPTETPDSHNTGG 804
 QY 841 SASP 900
 DB 805 STDGEMTEPGGSGGTAPDSGNTSGNTSGCKVRYSTTDWGS--FTGATINTTSP 861
 QY 901 SSVDLSTTVY-----WFTPD--GGSSTLYVNCMAAIGCNIIPASFGSVNPAIP 949
 DB 862 SPINGWTLAFTYPSGQITISVMSATOTLGRNVVLKNSGM-----NPTIP 906

KM municipal solid waste; glycoside hydrolase; avicelase.
 OS Aspergillus aculeatus.
 PN US2003108988-A1.
 PD 12-JUN-2003.
 PF 18-OCT-2002; 2002US-00155400.
 PR 28-JUL-2001; 2001US-00917376.
 PA (DING/) DING S.
 PA (ADNE/) ADNEY W S.
 PA (VINZ/) VINZANT T B.
 PA (HIMM/) HIMMEL M E.
 PI Ding S, Adney WS, Vinzant TB, Himmel ME;
 DR WPI; 2003-810853/76.
 XX
 PT New isolated thermal tolerant avicelase polynucleotide useful for
 PT detection of a polynucleotide encoding Aviii and for reducing cellulose
 PT in a starting material, e.g. municipal solid waste.
 XX
 PS Example 2; SEQ ID NO 7; 29pp; English.
 XX
 CC The invention relates to an isolated polynucleotide molecule encoding a
 CC thermostable Aviii polypeptide. The polynucleotide is useful for
 CC detection of a polynucleotide encoding Aviii. The polynucleotide is
 CC useful for reducing cellulose in a starting material which involves
 CC administering to the starting material, e.g. agricultural biomass or
 CC municipal solid waste, a polypeptide molecule of the polynucleotide. The
 CC method further comprises administering a second polypeptide molecule
 CC chosen from the glycoside hydrolase family of proteins. The present
 CC sequence represents the amino acid sequence of Aspergillus aculeatus
 CC avicelase III catalytic domain.
 CC
 SO Sequence 726 AA;
 QY
 Query Match 32.7%; Score 1680; DB 7; Length 726;
 Best Local Similarity 46.2%; Pred. No. 1e-80;
 Matches 346; Conservative 113; Mismatches 258; Indels 32; Gaps 17;
 QY 47 ATTQPTWNSVAL-GGGGFGVDFIVPNEGAPGILYRTDIGMYRMDANGRMIPLDWVG 105
 DB 1 AASQATYKNNVNTGGGGGFTPGIVFNPAGVAYARTDIGAYRLN-SDTWTPLMDWVG 59
 QY 106 ---MNMWGVNGVSIADPINTNKVMAAVGYTNSMDPNDGAILRSSDQATWQITPLPF 162
 DB 60 NDTWMDW---GIDALNTDPVDTDRVYVAVGYMTNMDPNVNGILRSTDDGDTWETKLPF 116
 QY 163 KLGANNPGRMGERLAVDNNNDILYFGAPSGGLMRSTDSGATWSQMTNFPDVGYTYIAN 222
 DB 117 KVGANNPGRMGERLAVDNNNDILYFGARSGGLMRSTDSGATWSQMTNFPDVGYTYIAN 176
 QY 223 PTDITGQSDIQGVVVAAPKSSSSIGQAKTIFVGVADPNPNPVFWSRDGATWQAVPGA 282
 DB 177 SSSST--YTDSPVGIAMVTFDSTGSSGSAATPRIFVGADAGKSVFSEKEDATAMWSGE 234
 QY 283 PT-GFIPKGVFPVNVHLYIATSNNGPYDGSAGDVWKFVSIGTTRISFVSTDTAN 341
 DB 235 POGYGLPHKGVLSPEETLYISYANGAGPYDGNNGYHKNITNSGWTDISP---TSLAS 291
 QY 342 DYFGYGLTIIDKHPNTIWAIOISWMDTIIIFRSTDGATWTRIMDWTYSNRLRYVL 401
 DB 292 TYFGYGLSVLDLQVPGTLMAALNCWMPDELIFRSTDSGATWSPIMWNGYPSINYYSY 351
 QY 402 DISAEPULTFGVGNP-PVSPPLGMDDEMAAIDPFNSDMLVGTGATLYATNDLTWKS 460
 DB 352 DISNAPWIDTSTDFPV---RVGMWVEALAIIDPFNSNMLVGTGTLVGGHDLTWDS 408
 QY 461 GGOIHIAFWKGLEETAVNDLISPPSGAPLISALDGLGFTADVTAVPTIFTSPTVFTT 520

RESULT 7
 ADD22927
 ID ADD22927 standard; protein; 726 AA.
 XX
 AC ADD22927;
 XX
 DT 15-JAN-2004 (first entry)
 XX
 DE Aspergillus aculeatus avicelase III catalytic domain.
 XX
 KM enzyme; Aviii; cellulose reduction; agricultural biomass;

XX	Xanthomonas campestris cellulase #1.
DE	
XX	Directed genetic engineering; galactomannanase; reduced activity;
KW	enhanced activity; xanthan gum production; suspension stability;
KM	emulsion stability; temperature resistance; pseudoplasticity; amylase;
KW	cellulase; extracellular protease; intracellular protease;
KM	glucose dehydrogenase; enzyme.
XX	
OS	Xanthomonas campestris.
PN	
XX	US2003036176-A1.
PN	
PD	20-FEB-2003.
XX	
PF	10-AUG-2001; 2001US-00927827.
XX	
PR	28-MAR-2001; 2001US-0279493P.
XX	
PA	(BOWE/) BOWER S G.
PA	(RAMS) RAMSEIER T M.
PI	Bower SG, Ramseier TM;
XX	
DR	WPI; 2003-625389/59.
DR	N-PSDB; ADD24893.
XX	
PT	New transformed cell or organism having reduced or enhanced activity of
PT	at least one protein, useful for producing xanthan gum, which are useful
PT	for providing formulations and properties.
XX	
PS	Claim 1; SEQ ID NO 47; 135pp; English.
XX	
CC	The present invention relates to polypeptide and polynucleotide sequences
CC	from Xanthomonas campestris which may be used for activity reduction or
CC	enhancement using directed genetic engineering. A transformed cell or
CC	organism having reduced or enhanced activity of at least one such protein
CC	e.g. galactomannanase can be generated by disrupting the gene encoding
CC	the protein. The activity of the protein is reduced by the presence of an
CC	antisense nucleic acid sequence. The nucleic acid sequence of the gene
CC	encoding the protein is a recombinant sequence having at least one
CC	mutation as compared to the wild-type gene encoding the protein. The
CC	transgenic cell or microorganism are useful for producing xanthan gum,
CC	which are useful for providing formulations and properties, such as long-
CC	term suspension and emulsion stability in alkaline, acid, and salt
CC	solutions, temperature resistance, and pseudoplasticity. The present
CC	sequence represents an enzyme relating to the present invention.
XX	
SQ	Sequence 555 AA;
	Query Match 16.5%; Score 848; DB 7; Length 555;
	Best Local Similarity 36.5%; Pred. No. 7.3e-37;
	Matches 193; Conservative 87; Mismatches 203; Indels 46; Gaps 14
Dy	23 TASFAVAALGVLPFAITASP-----AHATTQTPTYSNAVLGGGTVDSIVNEGAPGL 78
Db	TTTGAWLAVLSIL-LLFSTPEVRAAEPRTSGPYQWRSAVGCGFTVTGLFHPAERGIA 104
Oy	YVRVDIGMYNRDANNGRIIPLLDWGNMMWGNGVSIADPIPNNKMAVAVMGYNSM 138
Db	105 YARDVGGAYIKWDAAQQOATALTDMLGADNMIMLGIDAIVPADADALYLAAGTMYH- 163
Oy	139 DPNDGAIRSSDQAGTWOITTPLEPKLGGNMPGRGMGERLAVIDPNDNIILFYGAFSGKGI 198
Db	164 RAGTAVALRFSFRGRTPFERADLPFKLGQNLRRANGERLAVIDPHDGRVILLDSRIA-GLM 222
Oy	199 RSTDGATWSQMTNPEDVCTTYIANPLDTTGYSDIOGVYVAFFDKSSSSLGQASKTIYFG 258
Db	223 RSDRGAMAKVASPPDAALAGATRNHHGREQAV-GIAFVVFDAAISGNNGSPTPEIYVG 281
Oy	259 VADPNPFWMSDDGATQOAIVGCAPTGFIPIHGVPDPVNVHLVIATSNNGGPGYDSSGPV 318
Db	282 VSTEQTSLVSEEDACRSMAIEVAGQPERGLRPSHMAGSDGH-WYLSGDOPPGDLMAAGML 340

QY	319	WKFSVTSQTVTRISIPVSTQDANDYFGYSGLTIDROHNTIIVAAQISWDEPTIIFRSYD	378
Db	341	WKFTFPAQRMRWEISPIQ-PASGGGFGGAAVADPQHQVLLASFFRRRTTPDELYRSYD	399
QY	379	GGATWTRIMWTSYPNRSRLRYVLDISAEPWLTFCVQENPVPVPSPLGMMDEMAIDPENS	438
Db	400	GKKMTWPL-----LADAVFPHSAAPW-TAHATPH-----WMG-ALAIIDPFPG	439
QY	439	DRMLYXGATLTYATNDLTLMDSGGQIHIAF-----MVKGLEEFAVNDLISPSGAPL	490
Db	440	NHALFVETGYGIMASRNL-----ODFAAPQRPLOMWFQDRGLEETVPLDLISPMAGATL	492
QY	491	ISALGDIQGFTHADVTAVPSTIFTSVPVTTGTSVDYAEINPSIIVRAGS	539
Db	493	LSALGIDIDGFRHDDLDRV-QLYAGVFRITNGESIDDAAGAPQMVVRSST	540
RESULT 10			
ID	ABP73029	standard; protein; 1228 AA.	
XX	ABP73029;		
XX	03-JUN-2003	(first entry)	
DE	Amino acid sequence of the GuXA polypeptide.		
XX	GuXA, cellulase, glycoside hydrolase; enzyme; cellulose, sugar, biofuel,		
KM	detergent; pulp processing; paper processing; feed processing; textile.		
XX	Acidothermus cellulolyticus.		
OS	WO2003012109-A1.		
XX	13-FEB-2003.		
PD	28-JUL-2001; 2001WO-US023817.		
XX	28-JUL-2001; 2001WO-US023817.		
PR	(MIDE) MIDWEST RES INST.		
FA	Ding S, Adney WS, Vinzant TB, Himmel ME, Decker SR,		
XX	WPI: 2003-239526/23.		
DR	N-PSDB; ABZ77634.		
PT	Novel thermal tolerant GuXA polypeptide derived from Acidothermus		
PT	cellulolyticus, useful for reducing cellulose in a starting material, and		
PT	for the conversion of biomass to biofuels and biofuel additives.		
PS	Claim 3; Page 19; 47pp; English.		
XX	The present sequence represents a GuXA polypeptide. GuXA is thermostable		
CC	cellulase, and is a member of the glycoside hydrolase family of enzymes.		
CC	GuXA is useful for reducing cellulose in a starting material such as		
CC	agricultural biomass to sugars. This is useful in biofuel production.		
CC	GuXA is also useful in the conversion of biomass to biofuels and biofuel		
CC	additives, in detergents, pulp and paper processing, food and feed		
CC	processing, and in textile process. GuXA is also useful for raising		
CC	polycional and monoclonal antibodies that are useful in purifying GuXA,		
CC	or detecting GuXA polypeptide expression, as well as reagent tools for		
CC	characterizing the molecular actions of GuXA polypeptides		
XX	Sequence 1228 AA;		
QY	Query Match	16.0%; Score 823; DB 6; Length 1228;	
	Best Local Similarity	38.7%; Pred. No. 3.7e-35;	
	Matches 242; Conservative 44; Mismatches 189; Indels 150; Gaps 22		
QY	426	WMDEMAIDPENSRLYGTGATLTYATNDLTLMDSGGQIHIAFMVKGLEEFAVNDLISPP	485

Db 103 WMDRIALANGVNG-----GPGITTYLDAALSO-----QQGTTPEV---IEIYIYDL--PG 147
Qy 486 SGAPLISALGDIAGFTHADVTAVPSTIFTSPVFTTGTSDVAELNPSIIVRASFPSSSQ 545
Db 148 RDCAALASNGELPA-TAAGLQTY-ETQYIDPLASILSNKYSLSRLVTII-----EPDSL 200
Qy 546 PN---DRHVAISTDGGKQKMFQSGE-----PGCVTTGGTVAAASADGSRFTWAGDGG 593
Db 201 FNAVNTMSIQACATAVPYEEQGEIYALTKLHAIP-----WVIYIMDAHAHSGMT-GWPN 252
Qy 594 QPVVAVAFGNSMASSQGV-----PANAQIRSRVRNPKTFYALS 632
Db 253 NASGVOEVQKYLANSIGVNGIDGVTYNTANTPLKEPMTLTQQGQGVESANFYQNM 312
Qy 633 NGTFRSTGDTGVPFQVPAAGLPSGAVGVMFHAFVKEGDLMLAASSGLYHSTNGSSWS 692
Db 313 PDIDBADVAVDLYSRLLVAAGFPSS--IGMLIDTL-----RNGMGGRN 352
Qy 693 AITGVSSAVNV-----GFGKAPGSSYPAVF----- 718
Db 353 EPTGPSTATDVNTFVNQSKIDLRQHRGLMCNQAGAGLPQASPTDFPNAHLDAYWIK 412
Qy 719 VVGITIGVTGA-----YRSD-CGTTWVLINDQHQYGNMQA-----ITGDHANLRRV 766
Db 413 PPGESDGTSAASDPTTGKKSDPMCDPTTYT-----SYGVLTNALPNSPIAGQWPPAQFD 466
Qy 767 YIGTNGRGIVYGDIG-----APSGSPSPSVSPASPSLSPSPSSSPSPSP----- 814
Db 467 QLVANARAVPTSTSSPPPPSPSASPSPSPSPSPSPSPSPSPSPSPSPSPSPS 526
Qy 815 ---SPSSSPSSSP 871
Db 527 PSSP 586
Qy 872 GVKQYQKNDAPGNOIKPGLQVNTGSSSYDLSTVYTRYWFTPDGGSSTLYNCDMA 931
Db 587 GLKQYQKNDAPGNOIKPGLQVNTGSSSYDLSTVYTRYWFTPDGGSSTLYNCDMA 646
Qy 932 IGCGRIRASFGSVNPATPTADTYLQ 956
Db 647 MGCGRIRASFGSVNPATPTADTYLQ 671

RESULT 11

ABP73022 standard; protein; 762 AA.

XX AC ABP73022;
XX 03-JUN-2003 (first entry)
XX
XX Amino acid sequence of the Mana polypeptide.
XX Mana; mannanase A; glycoside hydrolase; enzyme; hemicellulose; sugar;
XX food; feed; paper pulp; biofuel; mannanase.
XX
XX Acidothermus cellulolyticus.
XX
XX WO2003012110-A1.
XX
XX 13-FEB-2003.
XX
XX 28-JUL-2001; 2001WO-US023819.
XX
XX 28-JUL-2001; 2001WO-US023819.
XX
XX (MIDE) MIDWEST RES INST.
XX
XX Ding S, Adney WS, Vinzant TB, Himmel ME;
XX WPI; 2003-248182/24.
XX DR N-PSDB; AB277633.
XX

PT Novel thermal tolerant mannanase A polypeptide derived from Acidothermus
PT cellulolyticus, useful for reducing hemicellulose in a starting material,
PT for processing of food, and as bulking agents in food stuffs.
XX
XX Claim 3; Page 18-19; 46pp; English.

CC The present sequence represents Mana, a thermostable mannanase A
CC polypeptide derived from Acidothermus cellulolyticus. Mana is a member of
CC the glycoside hydrolase family of enzymes. Mana is useful for reducing
CC hemicellulose in a starting material to simpler carbohydrate units, and
CC ultimately to sugars which are useful in the food, feed, paper pulp, and
CC biofuels industries. It is useful for the processing of food and in food
CC stuffs as bulking agents, and for the degradation of mannanase. Mana is
CC also useful to raise polyclonal and monoclonal antibodies that are useful
CC in purifying Mana, or detecting Mana polypeptide expression, and as well
CC as reagent tools for characterizing the molecular actions of Mana
CC polypeptides

SQ Sequence 762 AA;

Query Match 13.7%; Score 702; DB 6; Length 762;

Best local Similarity 35.0%; Pred. No. 5,4e-29;
Matches 214; Conservative 53; Mismatches 148; Indels 196; Gaps 27;

Qy 472 GLEETAVNDLISPPSGAPLISALGDIAGFTHADVTAVPSTIFTS-----PVFTTG 521
Db 2 GLVRRPARAFVAAAGT-AVAAAATLGSITMPATAPAFVYASGGQFVNLGLPYRYGG 60
Qy 522 TSVDYAE-----LNPSII-----VAGSPSPSQPNDRHVAISTGGKN 560
Db 61 TNNYLSYOSHADVDDVLAQAQMNLSVLRTWGPFIDIGSID-GSVP-----TIGGNKN 112
Qy 561 --WFGSPEPGCVTTG-----GTVAASADGSRFVMAFGDPQPVVYAVGF 602
Db 113 GPFYQWDP--STGAPAYNDGPTGLQGLDYALASAAHLKRI-----VVL 156
Qy 603 GNSMAAS-----QGVPAQAQIRSDRVNPKTFYALNGTFYR--STDGVTF-- 646
Db 157 TNDWKEFGMDQYDKMYGLPYHDFNYTD--PRTQAQYKQWVHLLNRVNSIGVTYKND 213
Qy 647 -----QPVAAQ--LPSSGA-----VGVMFYAVGKGGDLMLA-ASSGLYH 683
Db 214 PTFIAMELANEPKCVSGTLPTSGTCTOATIVMWQMSAYVKSIDPNHNVSGDEGFIY 273
Qy 684 STNGSSW---SAITGVSS-----AVNVGFGKAPGSSYPVAVGVTGAVRSDDC 725
Db 274 GSTQSGGMPYNDPSDGVNNALIRYKXIDFGTY--HLVFNW-----GQNAW 319
Qy 736 GTTWV---LIN-----DDQHQYGNWGAITGDHANLRRVYIGTNGR- 773
Db 320 GTQWIKQHIANAAGIKPTLIEFGWQTPRDSVYQTMQTVR-----TNGEA 367
Qy 774 GIVYDIDGAPSGSPSV-----SPASPSLSPSPSSSPSPSPSPSPSPSPSPSP 825
Db 368 GNMFMMLAGVNCQPYNDGFVNYPPSSGYATYLASEALALISTGTPSPSPSPSPSP 427
Qy 826 PSP 885
Db 428 PSPSPSPSPSP-----SASSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 471
Qy 886 DNOIKPGLQVNVNTGSSSYDLSTVYTRYWFTPDGGSSTLYNCDMAIIGCGRIRASFGSVN 945
Db 472 DNOIKPGLQVNVNTGSSSYDLSTVYTRYWFTPDGGSSTLYNCDMAIIGCGRIRASFGSVN 531
Qy 946 PATPTADTYLQ 956
Db 532 PATPTADTYLQ 542

RESULT 12

ABP73018 standard; peptide; 88 AA.

XX

AC ABB73018;
 XX 03-JUN-2003 (first entry)
 XX
 DE Amino acid sequence of AvilIII carbohydrate binding domain type II.
 XX
 XX Avicelase; AvilIII; glycoside hydrolase; enzyme; cellulase; biofuel;
 KM detergent; pulp processing; paper processing; feed processing; textile;
 KM cellulose.
 XX
 OS Acidothermus cellulolyticus.
 XX
 XX WO2003012090-A2.
 XX
 XX 13-FEB-2003.
 XX
 PD 28-JUL-2001; 2001WO-US023818.
 XX
 PF 28-JUL-2001; 2001WO-US023818.
 XX
 PR 28-JUL-2001; 2001WO-US023818.
 XX
 XX (MIDE) MIDWEST RES INST.
 PA
 XX Ding S, Adney WS, Vinzant TB, Himmel ME;
 PI
 XX WPI; 2003-248177/24.
 DR
 XX
 XX New thermostable AvilIII peptide from Acidothermus cellulolyticus, useful
 PT for degradation of cellulose or in generating anti-AvilIII antibodies for
 PT purifying recombinant AvilIII polypeptides from genetically engineered
 PT host cells.
 PS
 XX Claim 6; Page 8; 44pp; English.
 XX
 CC The present sequence is derived from a thermostable avicelase, designated
 CC AvilIII. AvilIII is a member of the glycoside hydrolase family of enzymes,
 CC and is a cellulase. AvilIII is useful in the conversion of biomass to
 CC biofuels and biofuel additives. It may be useful in the production of
 CC detergents, pulp and paper processing, food and feed processing and in
 CC textile processes. The thermostable AvilIII peptide is useful in the
 CC degradation of cellulose, and in generating specific anti-AvilIII
 CC antibodies that are useful in purifying recombinant AvilIII polypeptides
 CC from genetically engineered host cells, in detecting AvilIII polypeptide
 CC expression, as well as a reagent tool for characterizing the molecular
 CC actions of the polypeptide. The AvilIII polynucleotide is useful as a
 CC source of probes or primers in various diagnostic assays
 CC
 XX
 SQ Sequence 88 AA;
 Query Match 9.1%; Score 468; DB 6; Length 88;
 Best Local Similarity 100.0%; Pred. No. 1.1e-17;
 Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 869 VSGGVKQVQKNNDSAPGDNQIKFGLQVNTGSSVDLSTVTVRYWFTRDGSSSTLVNCD 928
 DB 1 VSGGVKQVQKNNDSAPGDNQIKFGLQVNTGSSVDLSTVTVRYWFTRDGSSSTLVNCD 60
 QY 929 WAAIGCGNIRASFGSVNPAFTADTYLQ 956
 DB 61 WAAIGCGNIRASFGSVNPAFTADTYLQ 88
 RESULT 13
 ABB73017
 ID ABB73017 standard; peptide; 89 AA.
 AC ABB73017;
 XX
 XX 03-JUN-2003 (first entry)
 DE Amino acid sequence of AvilIII carbohydrate binding domain type III.
 XX
 XX Avicelase; AvilIII; glycoside hydrolase; enzyme; cellulase; biofuel;
 KM detergent; pulp processing; paper processing; feed processing; textile;
 KM

KM cellulose.
 XX
 OS Acidothermus cellulolyticus.
 XX
 XX Key Location/Qualifiers
 FH Misc-difference 89 /note= "unspecified residue"
 FT
 XX
 XX WO2003012090-A2.
 XX
 PD 13-FEB-2003.
 XX
 XX 28-JUL-2001; 2001WO-US023818.
 XX
 PF 28-JUL-2001; 2001WO-US023818.
 XX
 PR 28-JUL-2001; 2001WO-US023818.
 XX
 XX (MIDE) MIDWEST RES INST.
 PA
 XX Ding S, Adney WS, Vinzant TB, Himmel ME;
 PI
 XX WPI; 2003-248177/24.
 DR
 XX
 XX New thermostable AvilIII peptide from Acidothermus cellulolyticus, useful
 PT for degradation of cellulose or in generating anti-AvilIII antibodies for
 PT purifying recombinant AvilIII polypeptides from genetically engineered
 PT host cells.
 PS
 XX Claim 6; Page 8; 44pp; English.
 XX
 CC The present sequence is derived from a thermostable avicelase, designated
 CC AvilIII. AvilIII is a member of the glycoside hydrolase family of enzymes,
 CC and is a cellulase. AvilIII is useful in the conversion of biomass to
 CC biofuels and biofuel additives. It may be useful in the production of
 CC detergents, pulp and paper processing, food and feed processing and in
 CC textile processes. The thermostable AvilIII peptide is useful in the
 CC degradation of cellulose, and in generating specific anti-AvilIII
 CC antibodies that are useful in purifying recombinant AvilIII polypeptides
 CC from genetically engineered host cells, in detecting AvilIII polypeptide
 CC expression, as well as a reagent tool for characterizing the molecular
 CC actions of the polypeptide. The AvilIII polynucleotide is useful as a
 CC source of probes or primers in various diagnostic assays
 CC
 XX
 SQ Sequence 89 AA;
 Query Match 9.1%; Score 468; DB 6; Length 89;
 Best Local Similarity 100.0%; Pred. No. 1.1e-17;
 Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 869 VSGGVKQVQKNNDSAPGDNQIKFGLQVNTGSSVDLSTVTVRYWFTRDGSSSTLVNCD 928
 DB 1 VSGGVKQVQKNNDSAPGDNQIKFGLQVNTGSSVDLSTVTVRYWFTRDGSSSTLVNCD 60
 QY 929 WAAIGCGNIRASFGSVNPAFTADTYLQ 956
 DB 61 WAAIGCGNIRASFGSVNPAFTADTYLQ 88
 RESULT 14
 ADD22925
 ID ADD22925 standard; protein; 88 AA.
 AC ADD22925;
 XX
 XX 15-JAN-2004 (first entry)
 DE Acidothermus cellulolyticus avicelase AvilIII CBD III #2.
 XX
 XX enzyme; AvilIII; cellulose reduction; agricultural biomass;
 KM municipal solid waste; glycoside hydrolase; avicelase.
 XX
 OS Acidothermus cellulolyticus.
 XX
 PN US2003108988-A1.

```
XX 12-JUN-2003.
PD 18-OCT-2002; 2002US-00155400.
XX 28-JUL-2001; 2001US-00917376.
XX (DING/) DING S.
PA (ADNE/) ADNEY W S.
PA (VINZ/) VINZANT T B.
PA (HIMM/) HIMMEL M E.
XX Ding S, Adney WS, Vinzant TB, Himmel ME;
PI WPI; 2003-810853/76.
XX New isolated thermal tolerant avicelase polynucleotide useful for
PT detection of a polynucleotide encoding AvilIII and for reducing cellulose
PT in a starting material, e.g. municipal solid waste.
XX Claim 16; SEQ ID NO 5; 29pp; English.
XX The invention relates to an isolated polynucleotide molecule encoding a
CC thermostable AvilIII polypeptide. The polynucleotide is useful for
CC detection of a polynucleotide encoding AvilIII. The polynucleotide is
CC useful for reducing cellulose in a starting material which involves
CC administering to the starting material, e.g. agricultural biomass or
CC municipal solid waste, a polypeptide molecule of the polynucleotide. The
CC method further comprises administering a second polypeptide molecule
CC chosen from the glycoside hydrolase family of proteins. The present
CC sequence represents the amino acid sequence of Acidothermus
CC cellulolyticus avicelase AvilIII CBD III #2.
XX
SQ Sequence 88 AA:
Query Match 9.0%; Score 462; DB 7; Length 88;
Best Local Similarity 97.7%; Pred. No. 2.3e-17;
Matches 86; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 869 VSGGVQYKNNDSAPGDNQIKRGLQVNTGSSVDLSITYVRYWFTRDGSSSTLYNCD 928
DB 1 VSGGVQYKNNDSAPGDNQIKRGLQVNTGSSVDLSITYVRYWFTRDGSSSTLYNCD 60
QY 929 WAIIGGNTIRASFGSVNPATPTADTYLQ 956
DB 61 WAIIGGNTIRASFGSVNPATPTADTYLQ 88
XX
RESULT 15
ADD22924
ID ADD22924 standard; protein; 89 AA.
XX
AC ADD22924;
XX
DT 15-JAN-2004 (first entry)
XX
DE Acidothermus cellulolyticus avicelase AvilIII CBD III #1.
XX enzyme; AvilIII; cellulose reduction; agricultural biomass;
KW municipal solid waste; glycoside hydrolase; avicelase.
XX
OS Acidothermus cellulolyticus.
XX
XX Key Location/Qualifiers
XX FT Misc-difference 89
XX FT /label= Unknown
XX
XX US2003108986-A1.
XX
XX 12-JUN-2003.
XX
XX 18-OCT-2002; 2002US-00155400.
XX
```

```
PR 28-JUL-2001; 2001US-00917376.
XX (DING/) DING S.
PA (ADNE/) ADNEY W S.
PA (VINZ/) VINZANT T B.
PA (HIMM/) HIMMEL M E.
XX Ding S, Adney WS, Vinzant TB, Himmel ME;
PI WPI; 2003-810853/76.
XX New isolated thermal tolerant avicelase polynucleotide useful for
PT detection of a polynucleotide encoding AvilIII and for reducing cellulose
PT in a starting material, e.g. municipal solid waste.
XX Claim 16; SEQ ID NO 4; 29pp; English.
XX The invention relates to an isolated polynucleotide molecule encoding a
CC thermostable AvilIII polypeptide. The polynucleotide is useful for
CC detection of a polynucleotide encoding AvilIII. The polynucleotide is
CC useful for reducing cellulose in a starting material which involves
CC administering to the starting material, e.g. agricultural biomass or
CC municipal solid waste, a polypeptide molecule of the polynucleotide. The
CC method further comprises administering a second polypeptide molecule
CC chosen from the glycoside hydrolase family of proteins. The present
CC sequence represents the amino acid sequence of Acidothermus
CC cellulolyticus avicelase AvilIII CBD III #1.
XX
SQ Sequence 89 AA:
Query Match 9.0%; Score 462; DB 7; Length 89;
Best Local Similarity 97.7%; Pred. No. 2.4e-17;
Matches 86; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 869 VSGGVQYKNNDSAPGDNQIKRGLQVNTGSSVDLSITYVRYWFTRDGSSSTLYNCD 928
DB 1 VSGGVQYKNNDSAPGDNQIKRGLQVNTGSSVDLSITYVRYWFTRDGSSSTLYNCD 60
QY 929 WAIIGGNTIRASFGSVNPATPTADTYLQ 956
DB 61 WAIIGGNTIRASFGSVNPATPTADTYLQ 88
XX
Search completed: May 14, 2004, 09:16:26
Job time : 70 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 14, 2004, 09:14:08 ; Search time 27 Seconds
(without alignments)
3409.455 Million cell updates/sec

Title: US-09-917-376-1

Perfect score: 5135

Sequence: 1 MDRSENIRLTMRSRRLVSL.....RASFGSVNPATPTADTYLQX 957

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2436	47.4	890	2 T35237	probably secreted
2	2037	39.7	839	2 D97013	probably secreted
3	1740	33.9	856	2 T00349	Avicelase III - As
4	1161	22.6	707	2 F72393	hypothetical prote
5	431.5	8.4	1779	2 T31085	xy lanase - Caldice
6	398.5	7.8	1711	2 T31337	1,4-beta-glucanase
7	397.5	7.7	1742	2 T37120	cellulase (EC 3.2.
8	372.5	7.3	473	2 S50755	hypothetical prote
9	346.5	6.7	901	2 A49227	slalidase - Actino
10	340.5	6.6	1331	2 A48954	mannan endo-1,4-be
11	308.5	6.0	611	2 S76211	hypothetical prote
12	301.5	5.9	1749	2 S75138	hypothetical prote
13	293	5.7	279	2 T10361	hypothetical prote
14	291	5.7	351	2 S50754	hypothetical prote
15	290.5	5.7	2232	2 T34434	hypothetical prote
16	290	5.6	1039	2 S02711	cellulase (EC 3.2.
17	288.5	5.6	1032	2 T34433	hypothetical prote
18	283.5	5.5	474	2 S15921	protein TPX-VT3 -
19	262	5.1	2468	2 A63412	hypothetical prote
20	261.5	5.1	1854	2 S36859	clpA protein - Clo
21	254.5	5.0	913	2 S20590	exo-alpha-sialidas
22	246	4.8	915	2 A43802	cellulase (EC 3.2.
23	245	4.8	879	2 A47704	endoglucanase I (E
24	244	4.8	1664	2 T18262	S-layer protein -
25	243.5	4.7	552	2 T08148	proline-rich myros
26	243	4.7	288	2 T17737	proline-rich prote
27	235.5	4.6	2124	2 A28452	proteoglycan core
28	229.5	4.5	496	2 T17908	proline/lysine-ric
29	227.5	4.4	2271	2 F90073	hypothetical prote

30	222.5	4.3	360	2 S12850	protein TPX - Ther
31	222.5	4.3	370	2 T45025	mycin MUC5B, trach
32	221	4.3	2132	1 A55182	aggreacan precursor
33	219	4.3	4776	2 E95206	cell wall surface
34	218.5	4.3	3164	1 WMBEH6	U36 protein - hum
35	217	4.2	241	2 AC2284	hypothetical prote
36	216	4.2	383	2 T46707	hypothetical prote
37	216	4.2	449	2 S16748	proteophosphoglyca
38	215.5	4.2	13055	2 T16580	proline-rich prote
39	215	4.2	339	2 T17636	hypothetical prote
40	215	4.2	13288	2 T03099	proline-rich prote
41	214	4.2	300	2 J02320	mycin, submaxillar
42	214	4.2	532	2 B35621	hydroxyproline-ric
43	211.5	4.1	1275	2 T33369	spore germination
44	211	4.1	602	2 AD2067	hypothetical prote
45	210	4.1	549	2 T17525	proline-rich prote

ALIGNMENTS

RESULT 1	
T35237	
Probable secreted cellulase - Streptomyces coelicolor	
C/Species: Streptomyces coelicolor	
C/Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999	
C/Accession: T35237	
R/Seeger, K.J.; Harris, D.; Parkhill, J.; Barrall, B.G.; Rajandream, M.A.	
submitted to the EMBL Data Library, September 1998	
A/Reference number: Z21572	
A/Accession: T35237	
A/Status: preliminary; translated from GB/EMBL/DBJ	
A/Molecule type: DNA	
A/Residues: 1-890 <SSE>	
A/Cross-references: EMBL:AL031515, PIDN:CMA20642.1, GSPDB:GN00070, SCODEB:SC5C7.30C	
A/Experimental source: strain A3(2)	
C/Genetics:	
A/Genes: SCODEB:SC5C7.30C	
Query Match	47.4%; Score 2436; DB 2; Length 890;
Best Local Similarity	48.6%; Pred. No. 1.7e-102;
Matches 475; Conservative 129; Mismatches 248; Indels 126; Gaps 19;	
QY	11 MRSRRVLSLAATASPAVAALGVLPATASPA--HAATTPYTWGNAIGGGGFPVDCI 68
DB	1 MRRTRITVTLAALAGLA---GSPPAASAEAPRAVAADSYTWGNATIDGGFVPGI 57
QY	69 VFNEGAPGLIYVETDIGMTRMDANGRWIPLLDWVGMMNGVGVSIADPINTNKW 128
DB	58 VFRTERKDLAVARTDIGAVRWQEEHWTPLLDHVQMDMGHTGVVALASDAVDPDRVY 117
QY	129 AAVGMYTNSWDPDGAILSSDGGATWQITPLPFGKGGNPGKMGKRLAVDPNDNIIY 188
DB	118 AAVGYTNDMPDPNPGVILRSADRGASVEKADLPFKGGNPGKMGKRLAVDPNDVLY 177
QY	189 FGAPSGKGLMRSDGATWSOMTNPDPVGYIAMPDTTGYOSPDIGVVWVAFDKSS--SS 247
DB	178 LGAPSGHGLMRSDAGVWSEVTAFFPVGNYADDPNDTSGYASDNGCITWTVDSTGGG 237
QY	248 LQGASRTIIFGVADPNNPVPFWSRDGATWQAVGAPGFTPHKGVDPVNVHLYIATSN 307
DB	238 AGTATRTIIVGVADKENVAYRSTDAATWRLAGQPTGYLAHKGVLDAENGVIYLAISDT 297
QY	308 GGPYDSSGDVWKFVSVTSGTWRISVPVSTDTANDYFGYSGLTIDRQHPVTIWAQISW 367
DB	298 GGPYDGGKRLYYATATGTWTDISPAABADT---YFGFGLTVDRQRPETWATAVYSSW 354
QY	368 WPTTITFRSTDGATWTRIMDWTSYPRSLRYLVDISAEWLTFFGVQPNPVPSPKLGW 427
DB	355 WPTQIFRSTDSGATWSQAWSYTSYDPRENRRTYMDVSSFWLTWGAPAPPEQTPKLGW 414
QY	428 DEAMADPFNSDMLTGTGATLTATNDLTKM--DSGGQIHTAPWKGLEETAVNDLISPPS 486

Db 415 TEALRIDPDSRMMYGTGATVYGTENTLTNMDDEGTPAIVEPMVRLGETAVNDLASPPS 474

Qy 487 GAPLISALGDIAGFTHADVTAVPSTIFTSVPFTTGTSDVYAEINLSIIVRAGSPFPSSOP 546

Db 475 GAPLISALGDIAGFTHADVTAVPSTIFTSVPFTTGTSDVYAEINLSIIVRAGSPFPSSOP 532

Qy 547 NDHVAEPTDGGKMWPGSGPVGTTGTVAASADGSRFVWAPGDPGQPVYAVGFGNSW 606

Db 533 ---HIAFTDNGAWMFGSTDPGSGVGGTVAAGADGSRFVWAPGDPGQPVYAVGFGNSW 587

Qy 607 AASGVANPANOIRSDRVNPKTFYALNSGTFRSTGGTGTQGVYAA-GLPSSAGVGMFLA 665

Db 588 QASGFLPAGALVEEDRVNPAFTFYGKSRFVSTGGATFTASATGTPAGD--GRRFYA 645

Qy 666 VPGEKEDMLAAS-----GLYHSTNGSSWGAITGVSAVWVGFKASGSSYPVAVFV 720

Db 646 LPGEKEDMLAASGADGPGYGLMSTDDGGTFTRLRGVDAADVYVGGKAPGASGYQLTFPS 705

Qy 721 GTTGGVTAAYSDGCTTGVTLINDQHOYGNMGOAITGDHANLRRVYIGTNGRGIVYGD 780

Db 706 AEIGGVNGIFRSTAGATWTRVNDADHGMGTGAITGDPVYGRVYATNGRGVYGD- 764

Qy 781 GGAEGSPSPSVSPSASLSPSPSSSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 840

Db 765 -----TSPTGGCTDPGPGPDPPTP----- 782

Qy 841 SASP 900

Db 783 -----TGACEVITYVTNQMPPGFO--ADVRLNTGT 811

Qy 901 S-----SVYLS-----TVTRVWFTF--DGSSTLYVNCW-----AAIGGNIRA--- 939

Db 812 SAMNGWLSLDMFPQGEVTRMNAHTQAGTSVTAARVNMNAGVAPGASVGGFTGSRSG 871

Qy 940 -----SFGSVNPAFTPA 951

Db 872 TNABEGFAVAGRACPTA 889

RESULT 2

D97013

probably secreted sialidase, several ASP-boxes and dockerin domain [imported] - Clostrid

C/Species: Clostridium acetobutylicum

C/Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001

C/Accession: D97013

R/Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

A/Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium C

A/Reference number: A96900; MUID:21359325; PMID:21359325

A/Accession: D97013

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-839 <KUD>

A/Cross-references: GB:AF001437; PIDN:AAK78895.1; PID:GL5023820; GSPDB:GN00168

A/Experimental source: Clostridium acetobutylicum ATCC824

C/Genetics:

A/Gene: CAC0919

Query Match 39.7%; Score 2037; DB 2; Length 839;

Best Local Similarity 49.5%; Pred. No. 1.4e-84;

Matches 392; Conservative 118; Mismatches 252; Indels 30; Gaps 16;

Qy 13 SRLVSLIAA-----TASPAVAALGVLFIAITASPAHAAT-TQPYTWSNVAIGGGFVD 66

Db 2 NKRIYSWVAGLSITFTGF-----VTHISAANKAAASVSGYKMNNAKIGAGGYVP 54

Qy 67 GIVNEGAPGILYVTRTDIGWYRMDAANGRWPLDWMVNMNMGVGVSIADAPINTNK 126

Db 55 AVIFKTEKDIYATDWMGAYRMDKANKWIPITD--GFSWYTMGCSISADPIDTNR 112

Qy 127 VMAAVGWTNSWMDPDGAILRSSDQATWQITPLPKLGGNMPGRMGRLAVDPNNNDI 186

Db 113 VYIAAGLYNMDODENAYIISDCKNTWKRYOLPRVGVGMGRNKGRLQIDPNDKI 172

Qy 187 LYEGAPSGKGLMRSTDSGATWSQMTNPPDVGYIANPTDTGYQSDIQGVVWAFDKSS 246

Db 173 LYIGASNGGLWMSBDYGCQWSKVDNFPDGTGVDPQNE--YTADKVGIVMEFFDST 230

Qy 247 SLGQAKTTFVGVAD--PNNPFWSRDQATWQVPAFTGFIPIHKGVFDPVNHVLYATS 305

Db 231 TKQSPQTMVYVGAADKTKGNNIYVINDGKTKWSAVKQPKGYLPHHGIL--ASDGLYISYS 289

Qy 306 NMGPPDSSGQVWKSIVSGTFRISPVSTDTANDYFVSGITTDROHPNTIMVATQ 365

Db 290 NMGPPDSSGQVWKSIVSGTFRISPVSTDTANDYFVSGITTDROHPNTIMVATQ 346

Qy 366 SWPDTIFRSTGATWTRIMDTSYNPSLRVYLDISAEMFLTFG--VQPNPVPSPKL 424

Db 347 RMPDEIRTSNAGTKMKTIMNMGVPRNLTGLNLDYSQPHLMDKRTGVTPEPDLVYL 406

Qy 425 GNMDMAIDPFNSDRMLYGTGATLYATNDLYKWSGGQIHLAPMKLEETAVNDLISP 484

Db 407 GNMWGLDIDPFNSDRMFTGTGATLYGTDLTWMDDKKNVDI SVKANGIECAVNDVVP 466

Qy 485 PEGAPLISALGDIAGFTHADVTAVPSTIFTSVPFTTGTSDVYAEINLSIIVRAGSPFPSS 544

Db 467 TKGAQLLSAVGDDCGFYHDDITVPSRMTTPNSATTSIDYAESVENFVVRGVNDVTSK 526

Qy 545 QPNDRIYVAFSTDGKMWFO--GSEPGVTTGTVAASADGSRFVWAPGDPGQPVYAVGFG 603

Db 527 NQDDKCCGISYDGGKMFSAKSNISGVYKAGTYAAGDACTIYWS--EKGANAISTDNG 585

Qy 604 NSWASQVAPANAQIRSDRVNPKTFYALNSGTFRSTGGVTF--QPYAAGLPSSGAVGM 662

Db 586 NKTTPSGGLPQAKVSDRNPCKFGFLNKFYIISTDAATFTQSSQGTPTGK--GI- 643

Qy 663 FPAVPEKEDMLA--ASSGLYHSTNGSSWGAITGVSAVWVGFKASGSSYPVAVFV 721

Db 644 FKTIVGHEGDIWLAGKGLMSTDDGGTFTRLRGVDAADVYVGGKAPGASGYQLTFPS 703

Qy 722 TGGVTAAYSDGCTTGVTLINDQHOYGNMGOAITGDHANLRRVYIGTNGRGIVYGD 781

Db 704 TIDGTAGIFRSDGATWTRVNDADHGMGTGAITGDPVYGRVYATNGRGVYGD- 763

Qy 782 GAPSGSPSPSVS 793

Db 764 GS-QPPTPESVT 774

RESULT 3

T00349

Avicelase III - Aspergillus aculeatus

C/Species: Aspergillus aculeatus

C/Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 16-Jul-1999

C/Accession: T00349

R/Araki, M.; Takada, G.; Kawaguchi, T.; Sumitani, J.

A/Title: The EMBL Data Library, June 1998

A/Description: Avicelase III from Aspergillus aculeatus.

A/Reference number: Z14141

A/Accession: T00349

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: mRNA

A/Residues: 1-856 <ARA>

A/Cross-references: EMBL:AB015511; NID:dl199887; PID:dl029971

C/Genetics:

A/Gene: avIII

C/Superfamily: fungal cellulose-binding domain homology

F/823-854/Domain: fungal cellulose-binding domain homology <PCB>

Query Match 33.9%; Score 1740; DB 2; Length 856;

Best Local Similarity 43.5%; Pred. No. 3.3e-71;

Matches 366; Conservative 139; Mismatches 237; Indels 40; Gaps 18;

Qy 30 AALGVPIAIIATASPAHAATQPYTWSNVAI--GGGGFVDGIVFNGAGIILVTRDIGMY 88

4 SSLALLCALGLKADAAASQAYTWKAVNTGGGGFTGCIYVNPBAKGVAVARTDGCAY 63
 89 RMDAANGRMILBLDMVG---WNNMGYNVGSIAADPINTNKWAAVGMTNSMNDPDAI 145
 64 RLN--SDDTWFTPLDMVGNDTWHDW---GIDALATPVDTRVYVAVGMTNEMDPRVSGI 119
 146 LRSSDQCATWQITPLPFLGGMPPGKMGGERLAVPNNNDIIFGAPSGKGLMRSTDGA 205
 120 LRSTYDQGTWETKLPFKVGNMPPGKMGGERLAVPNNKSIILYFAGRSQHGIMKSTDYA 179
 206 TWSQMTNPDPVGTIYANPTDITGYQSDIQGVVWVAFDSSSSSLGQASKTIIFGVADPNNP 265
 180 TMSVNTSTMTGTYTQDSSST--YTSDEPGIANTFDSSTGSSGSAITPPIFGVADAGS 237
 266 VFMSRDGATWQAVPGAPT-GFIPKGVFPDPVNHVLYATNSGTGPRYDSSGDVWKFVST 324
 238 VFKESEDAGATWAVSGEPQYGLPHKGYLSPEEKTLIYSYANGAPGVDGTNGVHKYNT 297
 325 SGVTRISFVSTDTANDYFGYSGLTIDRQHPNTIMVATQISWMPDTIIFRSTDDGATW 384
 298 SGVMTDISP---TSLASTYGYGGLVDLQVPGTLMVVALNCWMPDELFIRSTDSCATWS 354
 385 RIMWMTSYPNLSRLVYLISAEPMLTFGVQPNP--PVPEPKJGMDEAMAIDPFNSDRMLY 443
 355 PIMWMNGPSPINYYISYDISNAPWIDTSTQDFV---RVGMVWEALADIDPFDSNHWLY 411
 444 GTGATLYATNDLTWKDSGQIHIAPMVKLEETAVNDLISPPSGAPLISALDGLGFTTA 503
 412 GTGLTVGHHLTNMDSKHNTYKSLAVGIEEMATLGLITPBGPAALSAVDDDGFTYS 471
 504 DVTAVPSTIFTSPVFTGTSVDYAEALNPISIIVRAGSPDSSQPNDRHVAFTDGGKNMFQ 563
 472 DLDAAPNCAVHTPTGTNGIDYAGNKRPSNIIVRSASD--DYPT---LALSINFGSTWYA 526
 564 GSEBGGVTTGTVAAASADGSRFVWAPBGPQPVYVAVFGNSMAASQGVPAQAQIRSDAV 623
 527 DYAASTSTGTAVVALSADGDTVLMSSTGALVSKSQG---LTLVAVSSLPSCGAVIASDKS 583
 624 NPKTFYVLSNGTFRSTDGTGVTQFPAAGLPSSGAVGWVFAHPGEGDLMTLAASGLVH 683
 584 DNIVFYGSGAGAIYVSKNTATSTFKTVG--LGSSTTVNAI--RAHPSIAGDVMSTDKGLWH 641
 684 STNGSSWSAI--TGVSAAVNVGFGKAPGSSSYPAVFVGTIGGVAGYRSDCGTTVLI 742
 642 STDVSTFTQIGSGVTAAGSPFGKASSGTVVYVYGFETIDGAAGLPFSBDAAGTMQYI 701
 743 NDDHQYGNMGOA--ITGDHANLRVYIGTNGRGIIVYDGGAPSG-----SPSPSV 792
 702 SDASHGFQS--GSANVNVNDLQTYGRVFRGHERPHILRQSQREPAGRHGDGDGDTTSTKT 760
 793 SP 852
 761 STVSTVTLTKTTSSASTTSSSTTVKTTSSSTTSKASTTYYTKTTTSTTSSGTTTAA 820
 853 SS 854
 821 SA 822

RESULT 4
 F72393
 hypothetical protein TM0305 - Thermotoga maritima (strain MSB8)
 C/Species: Thermotoga maritima
 C/Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
 C/Accession: F72393
 R/Melton, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey
 Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
 C.M.
 Nature 399, 323-329, 1999
 A/Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
 A/Reference number: A72200; MUID:99287316; PMID:10360571
 A/Accession: F72393
 A/Status: preliminary

A/Molecule type: DNA
 A/Residues: 1-707 <ARN>
 A/Cross-references: GB:AE001712; GB:AE000512; NID:94980799; PIDN:AAD35393.1; PID:9498080;
 A/Experimental source: strain MSB8
 C/Genetics:
 A/Gene: TM0305

Query Match 22.6%; Score 1161; DB 2; Length 707;
 Best Local Similarity 35.2%; Pred. No. 2,9e-45;
 Matches 264; Conservative 123; Mismatches 275; Indels 89; Gaps 26;

52 YTMNSNAIAGGGGVVDGIVNEGARGILYVTRTDIGMYRWMAANGRMVPLIDWGMNMGY 111
 21 FEKSVLEINGGVFPGLIFHPASPGLLYATDVGLRWBETKRWQLPDLRQSDY 80
 112 NGVSIADPINTNKWAAVAVGMTNSMNDPDAIILSSDQCATWQITPLP---FKLGNN 168
 81 MGVLSVALBDSDDKRIYAMTKTQW--AGYGAIIILSEDEGETWTVNLQKYGIGN 139
 169 PGKMGGERLAVDNNNDIIFGAPSGKGLMRSTDGATWSQMTNPDPVGTIYANPTDITG 228
 140 DGRNAGERLQVDENFSSVLFMGT--TKYGLKKSDFGNKWKVDSFST----- 186
 229 YQSDIQGVWVAPADKSSSLGQASKTIIFGVADPNNPVFMSSRDGATWQAVPGAPTGTIP 288
 187 -----SVTFVLEDEKSGEKSPTPRIIFVGCSPKQ--IFTEDGITTWNLPLPMDLI 239
 289 HKGFDPVNHVLYATNSGTGPRYDSSGDVWKFVSTSGVTRISFVSTDTANDYFGYSG 348
 240 LRKQIH--DGILVYTSNALGPNGATRGAVMKYVIAQKKYDVTYPMKD-----FGYCG 291
 349 LTRDQHPNTIMVATQISWMPDTIIFRSTDDGATWTRIMDTSYPNKSLRYVLISAEPV 408
 292 IDVQ---NVVISTIDRWYPHDEIFLSLNGGFTWPLEKANP-----DINKAPV 339
 409 LTFGVQPNPVPSPKJGMDEAMAIDPFNSDRMLYGTGATLYATNDLTWKDSGQIHIAP 468
 340 IK---DLNP-----HWISD--VKIDPFDMNRALFTTGYGVWVTEYELKKSFEQ---MGK 384
 469 MYK-----GLEETAVNDLISPPSGAPLISALDGLGFTADVTAVPSTIFTSPVFTGT 522
 385 PVKMIENRGLLEETVVLQVLPPIGERPPLSALADWGGFRIESLDTPSSMY--PKLWMTSL 443
 523 SDVYAEALNPISIIVRAGSPDSSQPNDRHVAFTDGGKNMFQ--GSEBGGVTTG--TVAA 579
 444 GIAFAYONKSFVARNY---TYP---FLSYSDGGIINMEITVPEIGTIDGSRSLAAS 497
 580 ADGSRFVWAPGDDGQPVYVAVFGNSMAASQGVPA---NAQIRSDRVNPTFYAL--SNG 634
 498 NDGKTLVMSBAN--HEVIVASDGGKMKKALISVPVEFNYPPASDPVNPSEKFYIFPMKXG 555
 635 TFRSTDGTGVTQFPAAGLPSSGAVGMFHA---VPGKEGDIWLAA--SGLYHSTNGS 689
 556 DFLISDXGGSFEMK--GAKLPSFDNMWVSLSPFVLAPDRGDIWLLQNNGLRSDGGI 614
 690 SMSAIVGSAVNVGFGKAPGSSSYPAVFVGTIGTVGAYRSDCGTTVLIINDPOHOV 749
 615 TFERLGAVNDIAYVIGAPGPRGTDYPAIYANGVNVGYGIFMSTDEGKTMRIINDKQGF 674
 750 GNMWGAITGDHANLRVYIGTNGRGIIVYADI 780
 675 G-WIHIMIDMNEFRGILPLGTBERGIIIVEV 704

RESULT 5
 T31085
 xylanase - Caldicellulosiruptor sp.
 C/Species: Caldicellulosiruptor sp.
 C/Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000
 C/Accession: T31085
 R/Morris, D.D.; Gibbs, M.D.; Ford, M.; Thomas, J.; Bergquist, P.L.
 submitted to the EMBL Data Library, December 1997
 A/Description: Family 10 and 11 xylanase genes from Caldicellulosiruptor sp. Rte69B.1.

A;Reference number: Z20972
 A;Accession: T31085
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-1779 <MOR>
 A;Cross-references: EMBL:AF036924; NID:g2760905; PID:g2760909; PIDN:AA095326.1
 C;Genetics:
 A;Note: xymc

Query Match 8.4%; Score 431.5; DB 2; Length 1779;

Best Local Similarity 19.8%; Pred. No. 5.7e-12; Mismatches 382; Indels 477; Gaps 53;

Matches 252; Conservative 162; Mismatches 382; Indels 477; Gaps 53;

65 VDCIVFEGAGFIYV-----RTDIGMYRMDANGRMWIFLIDWVGMNMGVGVSIAD 120
 20 VGLIFHQEAKAAAYTVDFEGDTLSFFAYGKSNI-----VDMGN-AYNGKSIRRS 71
 121 PINTKVAVAAGMYTNSWDPNDGAILRSSDQATWQIT-----PLPFL-----G 165
 72 --NRSSIMDGVAV-----DVKNIMNNGTWWVSVYKHSYQKPAFGISAVYDDG 119
 166 GNMGRGGEELANDPNNNIIYFGAPSGKGLMRSTDSGATWSQMTNPFEDVGTIYANPTD 225
 120 SGVKSILGEVVAL-PN-----YWKI-----VAKMTPTNISN 150
 226 TTGQSDIGVYVWVAFDKSSSLGQASKTIIFVGVADPN---NPFWSR--DGGAT--WQ 277
 151 VRNLIIVHTIVESVDYNVD-----YQIMDDNSYLSNATYFSSGFESSGTTGEGQ 201
 278 A---VPGAPTFGFIPIHKGVFDPVNHVLYA--TSNTG--DYDG--SSGDVWKFSV-- 323
 202 ARGSGVTVKPDPSVAVYNGKYS---LVYSGRTSMWHGQIIPVDTLLEGQKYKISVWVY 256
 324 -TSGTWRISVPSIDTAND-YFGSGLTIDQHPPTINVAQISWPTTIFRSDGGA 381
 257 QNSGSTOKMSLTMRFRATDPSTYENILYNDVPSNTWVBSGSY-----SIPAGV 308
 382 TWRTIMDTSYPNRSLRY-----VLDIS--AEP----- 407
 309 TVSELILYVEQNANLAFWDDKITYLSKLAPEMEITSLIEKDYKRVGVVALSYKSI 368
 408 -----WLTGQVQNPVPVPSKLGW 426
 369 ASDTEKKMVLKHPNSITAGNEMKPELLISENNYNSKADDEVNFATSNINAIKHTLW 428
 427 MDEA-----NAIIP----- 435
 429 HEQTPDWEFFKDANGTLSKDALLSRLKOYITVVGKRYKGVYAMDVVNEAIDESQNGPR 488
 436 ----FNSDRMLYGTGATLYA-----TNDLTKMDSGQIHTAPWKGLBETAVNDL 481
 489 RSNMNYICGPEYIEKAFITAHBAPDAKLFYNDVNTENSQKQOFYNNMKSLE----- 542
 482 ISPSGAPLISALGLDGGFTHADVTAVPSTIFTSVFTT--GTSYDVAELNPSIIVRAGS 539
 543 ----KGVF-INGIG-LQCHINDWPSISIENTIKLFSIIPGLEIHTELDMSPYMGSS 596
 540 FDPSSQPD-----RHVAF-----STDGKKN---WQ 563
 597 TSVSTPRDLLIKQAMRYKELFDLFKYNVITNVTFWGLKDDYSMLSQNFGKSDYPLD 656
 564 GS-----EKGVTGTGTV-----AASA 580
 657 GNYSKAFAMGLIEFTVVPVNSTLPAAPAIQVPTPTSTPTPTPTVTSATPTPAPTASPA 716
 581 DGSFVWAPDPCQPVVAVGFGNSWMAASQGVAPANAQIRS-----DRVAP 625
 717 GGS--YWFPS-----SYGALKTWYANGMNSTTNVLPKIKIENVGTTAVDLSRVVY 767
 626 KTFYALNSGTFRSDGGVTFOP-----VAAGLPSSGA-----VGWFMHAVP 667
 768 RYWTYI--DGEAAQSVSVASSINPAYIDVYVVLGANAAGADYVVEVFGKSGAVILAAGS 826

QY 668 GKEGDMLAASGLYHSTNGSSWSA-----ITGVSSAVNVGCKKAPGSSYPVAVFV 719
 DB 827 YKEIRLSICKSSGSYNQSDNYSVRANSYIENEKVGYYIDVVLWGREPERNAQIKWYA 886
 QY 720 VGTIGGVTA---YNSDCGTT-----WYLINDQHQYGMWGAIGDHANLR 764
 DB 887 NGNLSSFTVNLNPKIKIENGTAVDLSRVKVRWYTTIDGEMQSVSVASSINPAYIDVR 946
 QY 765 RYVIGTNGRGIYV---GDIGGA-----PSGS----- 787
 DB 947 VYKLGANAGADYVVEVFGKSGAVILAAGSTKEIRLSICKSSGSYNQSDNYSVRANSY 1006
 QY 788 -----PSPSVSPASPSLSPSPSSSPSPSSSPSPSSSPSPSSPSP 828
 DB 1007 IENEKTYGIDVLVWKGKPSRGTKPAG--EYTPAPPTSTPTPTPTPTPTPTPTPTPT 1064
 QY 829 SP 883
 DB 1065 TVTATPT 1124
 QY 884 PGDNQIKPGLOVYNTSSSVYDLSYTVYVYFTRDGSSTLVYNCDMAAIGCNIRASFGS 943
 DB 1125 ASYGSIRPWFKIVNGSSSVYDLSRVKIRYVTVYDGDKPGSAV--CDMAQIGASVNTFRV 1183
 QY 944 VNPATPTADTYIQ 956
 DB 1184 LTRGVSADYYLE 1196

RESULT 6

T31337 1,4-beta-glucanase (EC 3.2.1.-) - Anaerocellum thermophilum (fragment)

C;Species: Anaerocellum thermophilum

C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000

C;Accession: T31337

R;Zverlov, V.; Mahr, S.; Riedel, K.; Bronnenmeier, K.

Microbiology 144, 457-465, 1998

A;Title: Properties and gene structure of a bifunctional cellulolytic enzyme (Cela) from omanis.

A;Reference number: Z21003; MID:96154434; PMID:9493303

A;Accession: T31337

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-1711 <ZVE>

A;Cross-references: EMBL:Z86105; NID:e1071329; PID:e350354; PIDN:CAB06786.1

C;Genetics:
 A;Gene: cela
 C;Keywords: glycosidase; hydrolase

Query Match 7.8%; Score 398.5; DB 2; Length 1711;

Best Local Similarity 22.9%; Pred. No. 1.7e-10; Mismatches 267; Indels 281; Gaps 40;

Matches 197; Conservative 116; Mismatches 267; Indels 281; Gaps 40;

262 PNNPVMSRSDGATWQA---VPGAPTFGFIPIHKGVDPVNVHLYIATSNTGPPYDGS--S 315
 25 PN-----WVNR--NMRGDSALKDQDNGLDLTGWMFADGHWKF---NLPMSTGTMLS 73
 QY 316 GDYWKFS---YISGTVTRISVPSTDTANDYFGYSGLTIDRQHPNTIMVATQI----- 365
 DB 74 WAYVEYKDAFVKSGOLEHI--LNQIEWVNDYF-----VKCHPSKKVYYVYQVVDGSKDH 124
 QY 366 SWW-----PDITIFSTDGAMTWTRIMWTSYPNRSLRVLILSAPMLTFGVOP 415
 DB 125 AMWGPALWQMERPSFEVYTSRGSSTVTE---TAASLAASIVLK-----DR 169
 QY 416 NPVPSPKLGWMDMAAIDPFNSDRMLYGTATLYATNDLTAKMDSGQIHTAPWKGLB 475
 DB 170 NPTKATYLOHAK-----LYBAEYTKSDAG-----Y 197
 QY 476 TAVNDLISPPSG-----APLISALGLDGGFTHADVTAVPSTIFTSVFTTGSVDY 526
 DB 198 TAANGYNSWSGFDYDELMAAVMLYLATNDSTYLTKEA-----SYQNWPFKISGSNTIDY 252

QY 527 -----AELNPSIIYRAGSPDSSQPNDRHVAFTDG-----GKNMFQ 563
DB 253 KWACWMDVHNGAAL--LAKITGKDIYKQIIESHLDYWTGNGERIKYTPKGLAMLD 309
QY 564 --GSEPGVTTGGTVAASAD-----GSRFWAPGDPGPVYAVGFG-- 603
DB 310 QMGSLRYATTTAFLAFVYSIDWVGCPSTKKEIYRKFGESQIDYALGSAGNS--FVVGFGTN 367
QY 604 -----NSWAASQGVPA-----NAQIRSD-- 621
DB 368 PPKRPHRTAHSSWADSGSIPSYHHTLYGALVGGPGSDSYTDDISYNNVEACDYNA 427
QY 622 -----RVNP---KTFYALSNGTFRSTGCVTPOVAAAGLPSSGAVGMFH 664
DB 428 GFVGALAMTYOLYGNPIPDFKAIETPTNDEFF-----VEAGINASGTNFIETK 476
QY 665 AVPGEGDMLAAS-----GLYHSTNGSSWSAITGVSAVNVFGKSAP----- 710
DB 477 AIVNNQSG-WPAKADKDKFRYFTVDLSBLIKAGYSPNQLT-LSTYNNQAKVSGPYVMDA 534
QY 711 --GSSYPAVFVVGIT--GVTGAYRSDCGTTWVLINDOHQYGNWGAITGDHANLRV 766
DB 535 SKNIYIILVDTGTILYVGGDKYKE-----VQFRIAPQNVGMNDNSFQDIKGV 588
QY 767 YIGT-----NGRGIYVDIGAPSGSPSPSASPSLSPPSSSPSSPSPSP 816
DB 589 SSGSVVTKTYIPLYGDVAKWGDGPGTSGATPTPTA-----TATPTPTPTPTPTPT 642
QY 817 SSSSSSSSP 875
DB 643 TS--TATPTPTPTPTPTPTPTPTPTPTA--TPSTPTPTPS-----TPSTPTPTPTPTPT 690
QY 876 QYKNNDSAPGNOIKPGQOVNVTGSSVDLSTVYRWVFTPRDGSSTLYVNCDMAIIGCG 935
DB 691 LYAKETNSTNTIRPMLKVNVTGSSSIDLSKVTIRYVTVVGDRAQSAIS-DMAQIGAS 749
QY 936 NIRASFGSVNATPTPTADTYLQ 956
DB 750 NVTPEFVTLSSSVSGADYYLE 770

RESULT 7

117120
cellulase (EC 3.2.1.1-) precursor, thermoactive - Caldocellum saccharolyticum
C:Species: Caldocellum saccharolyticum
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Nov-2000
C:Accession: 117120; A43745
R:Te'o, V.S.; Saul, D.J.; Bergquist, P.L.
Appl. Microbiol. Biotechnol. 43, 291-296, 1995
A:Title: CelA, another gene coding for a multidomain cellulase from the extreme thermoph
A:Reference number: Z18658; MUID:55336703; PMID:7612247
A:Accession: 117120
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1742 <TEO>
A:Cross-references: EMBL:L32742; NID:G537499; PID:G537500; PIDN:AAA91086.1
R:Lueethi, E.; Bhana Jasmat, N.; Grayling, R.A.; Love, D.R.; Bergquist, P.L.
Appl. Environ. Microbiol. 57, 694-700, 1991
A:Title: Cloning, sequence analysis, and expression in Escherichia coli of a gene coding
A:Reference number: A43745; MUID:91247819; PMID:2039230
A:Accession: A43745
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1516-1544, 'A', 1546-1742 <LUE>
A:Cross-references: EMBL:M36063; NID:g144292; PIDN:AAA72860.1; PID:g144293
C:Genetics:
A:Gene: celA
C:Keywords: glycosidase; hydrolase

Query Match 7.7%; Score 397.5; DB 2; Length 1742;
Best Local Similarity 21.7%; Pred. No. 1.9e-10;
Matches 189; Conservative 109; Mismatches 269; Indels 303; Gaps 35;

QY 262 PNNPVMRSRDGATMQA-----VPGAPTGRIPIHGVDFPVNHLV-LATSNTGPGYDSSG 316
DB 48 PN-----WVRN--NMGDSALKDQDNGDLTGCGWPDAGXFKNLEMSYTGMLSWAAY 100
QY 317 DVKKFSVTSCTWTRIPVPSTDPANDYFGYGLTIDROHNTIMVATQI-----SMW- 368
DB 101 EYDADVKSQGLEHT--LNGIEVNDYF-----VKCHSKYVYVQVGDGKDHAMWG 151
QY 369 -----PDTIIFSTDGATWTRIMWTSYPNRSLRYVLDISAPMLTFGVQPNPV 419
DB 152 PAEWQMBERSFVKTQSSPSGA-----VVAETASLAASIVLDRN 193
QY 420 PSFKLGMDNMAAIDPNRPMLYGTGATLYATNDLTKMDSGGQIHIAPMVKILEEYAN 479
DB 194 PTYAAATYLOHA-----KDIYERAEVYKSSG-----YTAAN 224
QY 480 DLSPSPG-----APLISALGDLGTFHADVNAVPSITFTSPVFTGTSDY----- 526
DB 225 GYNSWSGFYDELSTMAAWLYLATNDSTYLTKE--SYQNPWKISGSNIIDYKMAH 279
QY 527 -----AELNPSIIYRAGSPDSSQPNDRHVAFTDG-----GKNW----- 561
DB 280 CWDVHNGAAL--LAKITGKDIYKQIIESHLDYWTGNGERIKYTPKGLAMLDQWGS 336
QY 562 -----FGSEPGVTTGGTVAASADG--SRFWAPGDPGPVYAVGFG----- 603
DB 337 LRYATTTAFLAFYYSWSGCGPTGKKEIYRKFGESQIDYALGSTGRS--FVVGGTNPKR 394
QY 604 -----NSWAASQGVPA-----NAQIRSD----- 621
DB 395 PHRTAHSSWADSGSIPSYHHTLYGALVGGPGSDSYTDDISYNNVEACDYNAFVG 454
QY 622 -----RVNP---KTFYALSNGTFRSTGCVTPOVAAAGLPSSGAVGMFHAVPG 668
DB 455 ALAKMYLVGNGPIPDFKAIETPTNDEFF-----VEAGINASGTNFIETKAL-- 501
QY 669 KEGDLWLAASGLYHSTNGSSWSA-----ITGVSAVNVFGKSAPESSYPAYF 718
DB 502 -----VNNQSGWPARATNKLKFRYFTVDLSBLIKAGYS-----PNQL 537
QY 719 VVGIT--GVTGAYRSDCGTTWVLIND-----DOHQYGNWGAITGDHANLR 765
DB 538 TLSTYNNQAKVSGPYVWSSRNYYIILVDTGTLYVPGQDKYKKEVQFRIAPQVQ-- 596
QY 766 VYIGTNGRGIYVDIGAPSGS-----PSPSPSPSPSPSPSPSPSPSPSPSP 807
DB 597 ---WDSNDYSFQDINGVSSGVKTKIYPLYEDIKWGBEPGTS-----GVSPPTAS 648
QY 808 SPSPPSP 867
DB 649 VTFPTPTPTPTA-----TPPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 700
QY 868 P-VSGGVKQYKNDAPGDNQIKPGQOVNVTGSSVDLSTVYRWVFTPRDGSSTLYVN 926
DB 701 PATSGQIKVLYAKETNSTNTIRPMLKVNSSSIDLSRVIRYVTVVGDRAQSAIS 760
QY 927 CDMAAIGCGIRASFGSVNATPTPTADTYLQ 956
DB 761 -DMAQIGASNVTFKFKVTLSSSVSGADYYLE 789

RESULT 8

S50755
hypothetical protein VSP-3 - Chlamydomonas reinhardtii
C:Species: Chlamydomonas reinhardtii
C>Date: 14-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 21-Jul-2000
C:Accession: S50755
R:Moessner, J.P.; Molendijk, A.J.; van Egmond, P.; Kils, F.M.; Goodenough, U.W.; Harting,
Plant Mol. Biol. 26, 947-960, 1994
A:Title: Domain conservation in several volvoclean cell wall proteins.
A:Reference number: S50754; MUID:95093034; PMID:8000007
A:Accession: S50755
A:Status: preliminary

A:Molecule type: mRNA
A:Residues: 1-473 <MOB>
A:Cross-references: EMBL:L29029; NID:G530875; PIRID:AA53953.1; PID:G530876

Query Match 7.3%; Score 372.5; DB 2; Length 473;
Best Local Similarity 33.0%; Pred. No. 5.4e-10;

Matches 132; Conservative 36; Mismatches 139; Indels 93; Gaps 17;

```
QY 492 SALDGLGFTADVAVPSTIFTSFVFTGTGVDVAELNPSIIIVAGSFDPSQPN----- 547
DB 30 SAAGPYTYTYANSOVIASTLRLSLV---ETWLKTI--LTPGLAKNPIQTTSQPSAAL 84
QY 548 ---DRHVAFTDGGKMW---FGSEPGVTTGTGTVAAASADGSRFYWAPDPCGPVTVYAVG 601
DB 85 AGADAVYVILSNRYRYWYTEKMGSSPNKY-----RDLKDWNVAGSS---LVLIDG 131
QY 602 FGNWMAASQGVPAANAQISDRVNPTFTALNSG--TFPRSTGCGVTFQVVAAGLPSGAV 659
DB 132 YSTASGTVTFQQLDAVLTGKAGSGCTGALYNGNVNVVRANSSSPFGKITPSLTVKGS- 190
QY 660 GWMFHAVPGKGDLLAASGLYHSTNGSSMSAITGVSAVNVGPGKSA---PGSSYP 716
DB 191 -----SRGSGLTGCTSGAVLFSSNPDKMTKAVT--ASALTWSYVKGALITFTGSSPAM 241
QY 717 VFVVGTTGVTGAVYRSDCGTTWVLINDQHQYGNWGAQITGDHANLREVTYGTNGRGIV 776
DB 242 PHLKG-----YEDMS-----GVAVT--LANFNT----- 263
QY 777 YGDIGGAPS---GSPSPSPSPASPSLSPPSPSSSPSPSPSPSPSPSPSPSPSPSPSPSP 830
DB 264 ---GASPSPPSPKASPSFVSPSPSPKASPSPSPPKASPSPSPPKASPSPSPPKASPS 319
QY 831 SPSPSRSPSPASPSPSPPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 868
DB 320 SPSPSPPSPKASPSPSPPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 359
```

RESULT 9

A49227
A:Accession: A49227
A:Title: Complete nucleotide sequence of the Actinomyces viscosus T14V sialidase gene: F
C:Species: Actinomyces viscosus
C:Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 22-Oct-1999
C:Accession: A49227
R:Yeung, M.K.
A:Reference number: A49227; MUID:93114861; PMID:8418033
A:Contents: T14V
A:Accession: A49227
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-901 <YEU>
A:Cross-references: GB:L06898; NID:G289087; PIDN:AA21932.1; PID:G141652
A>Note: sequence extracted from NCBI backbone (NCBIN:121598, NCBIP:121599)

Query Match 6.7%; Score 346.5; DB 2; Length 901;
Best Local Similarity 22.0%; Pred. No. 1.7e-08;
Matches 228; Conservative 99; Mismatches 304; Indels 405; Gaps 51;

```
QY 30 AALGVLPPIAIT-----ASPAHAATQPYTWSNVAICGGFVDGIVNEGAPGI-LYVRTD 83
DB 14 ALTGLSLPLAATGLIAAPPAHAVPT-----SDGLADVTITQVANAAPAGLXSVGD 62
QY 84 I-----GMYRMDANGRWIPLLDWVGNNWNGVVSIAADPINTNKWAAVAG 132
DB 63 VMTFNITLTNTSGEASHYAPASTNLSGNVSKCRWR-----VPAGITKT-DCTG 110
QY 133 MYTNSWDNDGAILASSDGGATWQITPLPFKCGNNPGKMGKRLAVDPNNNDIILYFGAP 192
DB 111 L-----ATHVYAECLKKGGFTQJAY-EVKAVE-----Y 139
QY 193 SGKGL-WRSTDSGATWSQMTNFPDQVTIANPT-----DTTGYSODIGVVWVAFDKS 244
```

DB 140 AGKALSTPETIKATSPVKANSILRVESITPSSSCENTKGLDVTYTVRVASV----- 191

QY 245 SSSLSQASAKTI FVGVA-----DPNNPVFMSRDGATWQAVPGATGFIPIHKGFV---DPV 296

DB 192 -----SDKTIINVAALRESSFDLGRQCHNG-----GLKPKGKAVYNNKPL 230

QY 297 NHVLYATSNSTGPPYDGGSGDVWKFVSVT-----SGTWRIISP 333

DB 231 THITITQADVAGR-----WTPSITLLATGTDGATLQTLTATGNPINVGDHPQATP 281

QY 334 VPSTD-----TANDYFSGSLTI-----DR 353

DB 282 AAPADASTELPASMQAOLHAANTATDNYRIPAITTAPNGDILLISYDERPKDNGSGDA 341

QY 354 QHNTIMVATQISWMPDITIFRSTGCGATW-----TRIMWTSYPSRLRYVDIS 404

DB 342 PNNNHVOR-----RSTDGKTSAPFTYHGTETGKKGYSPPS--YVVDHQ 387

QY 405 AEFMLTFGVQPNPVPSPKLGWMDMAIDPFN-----SDRMLYGTGATLYATN 453

DB 388 TGTIFNFHVX-----SYDGWGSRGCTDPENRGIIQAEVSTSDNGMTWTRHTI--TA 439

QY 454 DLTK---WDS-----GGQIHIAPMYKGLERTAVNDLISPPSGA-PLIGALDGLGFTH 502

DB 440 DITKDRPWTARFASGGQIDIGHFAGRLVOQYT--ITTAGAQAQAVSVSDHDKTW 496

QY 503 ADVTAVPSTIFTSFVFTGTGVDVAELNPSIIIVAGSFDPSQPN-----RHVAFSTDG 558

DB 497 QACTPI-----GGMDE--NKVVELSDGSLMNSASDGSGRKRAHSTDDG 541

QY 559 KNMFQSEPGVTTGTGTVAAASADGSRFV-----WAPDPCQPVYVAVFGNSWMAASQVP 613

DB 542 QTM---SEP---VSDKNLPDSVQNAQIRAFPMNAAPDPRAKVL----- 579

QY 614 ANAQISDRVNPTFTALNSGTFTYRSTGCVTQPVAAGLPSSGANGWFMHAVPGKGD 673

DB 580 ---LHSHSPNRP--WSRDRGTISMSCDDGAS----- 606

QY 674 WLAASGLYHSTNGSSMSAITGVSAVNVGPGKSAFGSSYPVAVV--GTIGVGTGAYR 731

DB 607 W--TTKRVFHE-----PFGVYTTIAVSDSITG----- 632

QY 732 SDDCGTTWVLINDQH--QYGN-WGQAITGDHANLRRVYIGTNGRGIVYGDIGGAPSGS 787

DB 633 -----LISDEAHNGALYGGIWKYKNT-----NHWLGEQCGKRAEBSPA 671

QY 788 PPSPSVSPASPSLSPPSPS--PSSSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 841

DB 672 PSPTAAPSAAPTEKPAAPSAPSAAPTQAPAPSSAPSPSAAPSPSAAPSPSAAPSP 731

QY 842 ASP 893

DB 732 PAPAAPSAPQOTGPTAAPAPETTSAPAPAPPTQAPVAPSEVPTQAPQAPSSAPKPGAT 791

QY 894 ----QVNV---TGSSS 902

DB 792 GRAPSVNPKATGAT 807

RESULT 10

A48954
A:Accession: A48954
A:Title: The beta-mannanase from "Caldocellum saccharolyticum" is part of a multidomain
C:Species: Caldocellum saccharolyticum
C:Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 08-Oct-1999
C:Accession: A48954; B43745
R:Gibbs, M.D.; Saul, D.J.; Inub, E.; Bergquist, P.L.
A:Reference number: A48954; MUID:93119139; PMID:1476429
A:Status: preliminary
A:Molecule type: nucleic acid

Query Match	6.6%;	Score 340.5;	DB 2;	Length 1331;
Best Local Similarity	22.7%;	Pred. No. 5e-08;		
Matches 140;	Conservative 97;	Mismatches 189;	Indels 191;	Gaps 29

[illegible]

O. K.; Okumura S.; Shimpō, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
A:Accession: S76211
A:Status: nucleic acid sequence not shown; translation not shown
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*
B:
A:Reference number: S74322; MWID:97061201; PMID:8905231
A:Accession: S76211
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-611 <KAN>
A:Cross-references: EMBL:D90914; GB:AB001339; NID:g1653477; PIDN:BAAL8470.1; PID:d1019207
A>Note: The nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 6.0%; Score 308.5; DB 2; Length 611;
Best local similarity 26.7%; Pred. No. 5.5e-07;
Matches 142; Conservative 51; Mismatches 180; Indels 159; Gaps 21;

Dy 429 EAMA-IDPFNSDR-----MLYGTATLYATNDITKWDGQIHIAPVKGLSEETVANDLI 482
Db 151 EASAGQVEIRINSSSPRIYSKYGVGEENKGKIPQ-----GGVIRLVAVEGSSKGAIDAV 206
Dy 483 S-----PPSGAPLISALG-----DLGGFTADVAIVESTITTSPIYTTGTSVDY 526
Db 207 TRIMADIPISKRGSTFGSGSDTPPALMISEGVDKKCTAGPDVTGTPPAVPKVG----- 261
Dy 527 AEINPSIIVR-AGSP-PSSQPMDRHVAFSTDGSKWMFOGSEPQCVTGTGVVASADGSR 584
Db 262 -EIRGNIVLAGCGSTDLTDYIIDLNLSIQNAGGFAAVRASTPMPSIPLSALSARSQNR 320
Dy 585 FVMARGPDPQPVVYVVGFGNS-----WMAQGVPANAQIRSDRVNPXT 627
Db 321 -----VYTLNVGNNDNISLP.RSGEGQPVDDGYYYESSSLDKRFELPNPTNGGKK 368
Dy 628 FYALNSNGTFYRSTDGCVT-----FQPV--AAGLPSSGAVGWFAHPVKREGDLTLAAS 678
Db 369 YRLVYTGGIGRNAEAINGTCNTADGCCQPTDIQIFGLSSSGSI----- 409
Dy 679 SGLHSTNGCGSSWSAITGVSSAVNVGFSGKAPSSSYAVFVGNITGTVGAYRSDDCGTT 738
Db 410 ----CMGNSSRTTKAFV-LAPNYDIGRTGNS----YEGMLLCKT-YAKGNCAV- 454
Dy 739 WYLINDQHQQYNWCQAITGDHANLRVYIGTGRCIGVGDIGAPSGSPSVSPASP 798
Db 455 -----NNGKLAAVIQT-----SQTWGSVPPDL-RPILTAHSPTLTSSSW 491
Dy 799 S-----LSPSPSPSSSPSPSPSSSPSSSPSSSPSPSPSPSPSPSR 836
Db 492 SQIDGSLADDDAITEPPIYTGSGFLTRASSSPSPSPSPSPSPSPSPSPSPSPSPSPSP 551
Dy 837 SPSPSPASSPSPSPSPSPSSSPSSSPSTPSSSPS-----GVKQYQKN 880
Db 552 SPSPSPSPSPSPSPSPSPSPSPSPSPSPSPPTPVTVNQNKACDLDGTYSGQGN 603

RESULT 12
S75138
hypothetical protein slr1753 - *Synechocystis* sp. (strain PCC 6803)
C:Species: *Synechocystis* sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
C:Accession: S75138
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
O. K.; Okumura, S.; Shimpō, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda,
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*
B:
A:Reference number: S74322; MWID:97061201; PMID:8905231
A:Accession: S75138
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1749 <KAN>
A:Cross-references: EMBL:D90903; GB:AB001339; NID:g1652127; PIDN:BAAL7052.1; PID:d101785
A>Note: The nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Genetics:

A:Start codon: GTG

Query Match 5.7%; Score 301.5; DB 2; Length 1749;
 Best Local Similarity 19.0%; Pred. No. 3.9e-06;
 Matches 225; Conservative 117; Mismatches 402; Indels 439; Gaps 46;

QY 17 VSLAATRSFVAALGVLPITATSPHAATTPYTMNSNAIGG----- 62
 DB 23 LALLACSSSFSG-----NVALQNIPTAPDGTGTVAQGNQFNIGGGSLSGDCQNLPHSLQ 78
 QY 63 -----GFVDDIVFEGAPGLIYVTTDQGMWRMDA 92
 DB 79 QFGLDQGOIANFLSNPDIRNLITRIVGSDASTIINGLIQVSGNANLFLMNPAGMIFGPNA 138
 QY 93 A-----NGRWIPLDMVGNNMGNGVNS----- 116
 DB 139 SINPGRPVVTGSAIGFNDQMPV-----PSDDYNALIGNPQGFADLANPGLITNA 193
 QY 117 -----IAADPINTNKVAAVGMTNSMDPNDGAILRSS----- 149
 DB 194 GDLSTEGKNTLPAGNIIVTGSILAAPGNITVAAPGQNRIRISQAGSLILEVEVSPQ 253
 QY 150 -DQATWQITPLPFLTGGMPGRGKGEHLAANDPND-----NIIYFGAPSGKGLMRSD 202
 DB 254 MNOGSGSEVLDPLTLTGASNLDG--LAQPNQSVTTNGTNAIVSPPLGSVTISGNDV 311
 QY 203 SGATWSQMTNFPDVGTYIANPTDTGYOS--DIQ-----GVVWVAFD-KSSSSLGQASK 253
 DB 312 AS---GKSTNISS--GGQVAIAGDQIAVQCATVDSNGGCGIVRIGDQCGQTLTPNAQ 367
 QY 254 TIFVGVADPNNPV-----FMSRDGA--TWOAVPAGPTGFIPIKGVDPVNHVLYATSN 307
 DB 368 TLI---DSNGVVKADALLTGNGGIVIWADSTRPSNI-----SAG 406
 QY 308 GGPVDSGGVWKSVTSGTWRISPVSTDTANDYFGISGLTIRQHPNTIMATQISW 367
 DB 407 GGTGNGNG--FVETSG-----AKSLWDDTARVNTATMELDT 444
 QY 368 W---PDITIFSTOGATWTRIMD---WTSYPNRSILRVLDISAPMLTFGQPPPP 420
 DB 445 WLDPLEIIVGTDLADPKIVSVLTITTSLDNGNVIIQADQSLAVQNFADPSAP-- 502
 QY 421 SPKLGMDMAIDPFNSDRMLYGTGATLYA-----TNDLTMDSGGQIHI-- 466
 DB 503 -GNLTFPSPTITIDLFS-----LGTGSIIFANTGPIINTGNTLVTSPTNLDPDNKIQUNA 557
 QY 467 ----APVKGLSEETAVN--DLISPPS-----GAPLISALGDG----- 498
 DB 558 NTTFAPGYDIYFRKSVNGGFLLGNANFVFPDDGAGITTPKSFVATATEIYVNDIYT 617
 QY 499 -----GFTHADVT-AVPSITIFPSVPTTG-----TSVDYALNP 531
 DB 618 QGNQIFDGVFGLQPNLITSSAGSVIFINNILLNSGLQVQTAQNIIVSQSSLSAVALIS 677
 QY 532 SIIVRAGSPDSSQPNDRHVAF--STDGKNMFGQSEFGVTTG-----GTVA 577
 DB 678 DVLNAG-----QNVSRGNINTRGANDIQAL--GNISTGSIVSPRGKNGNVI 725
 QY 578 ASADG--SRPVMAGDPQGPVVAVVGFSNA-----ASQGV 612
 DB 726 LNAAGTLTGYIETSGTNGGDVTTSSGNTSTAYIDTRFGDGLAIDSLGAVSIESKGD 785
 QY 613 PANQIRSDRNPRTFYALSNKTPFRSTDGVTTFQVAAGLBSSGAVGMFHAHPKE-- 670
 DB 786 ITTAFIDTGVASIESFNEGTVGNVFLADGSITTYNYITAGKNGG--DIFQGESIEII 843
 QY 671 -----GDWLIAA-----SSGLYHSTNGS-----SMAITGV 697
 DB 844 DYLNTYSGQSDVYVEAPLDISISYITGGGEGKGNFLQAGDITTSYIDTSANGG 903
 QY 698 SSAVNVG-----FGKAPSSSYPAVFVGT 722
 DB 904 DIFIGSGDTEVGYLFTKGYEGRGDDVYVETGRYPRAIDGILLGEBGFPSSVYTAGLTVG-- 962

QY 723 IGGVTGAYNSDDCCTTWLINDQHQYGMKQAITGDHANLRVYICTNGR--GIYVGD 780
 DB 963 -----GSVYI-----QFGSGEPPIINPI-----TNGTIGALISGDD 994
 QY 781 GQAPSGSP-----SPSVSPASLSLSPSPSSSPSPSPSPSSSPSSSPSS 825
 DB 995 NTVPICITPTPTDITLNDITITTEPPEPPEPPEPPEPPEPPEPPEPPEPPEPPE 1054
 QY 826 PSP 868
 DB 1055 PEPPEPPEPPEPPEPPEPPEPPEPPEPPEPPEPPEPPEPPEPPEPPEPPEPPEP 1097

RESULT 13

110361
 hypothetical protein 92 - Oryza pseudosugata nuclear polyhedrosis virus
 C/Species: Oryza pseudosugata nuclear polyhedrosis virus, OpMV
 C/Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
 C/Accession: T10361
 R/Authors: C.A.; Russell, R.R.; Funk, C.J.; Evans, J.; Harwood, S.; Rohmann, G.F.
 Virology 229, 381-399, 1997
 A/Title: The sequence of the Oryza pseudosugata multicapsid nuclear polyhedrosis
 A/Reference number: Z17011; PMID:97271300; PMID:9126251
 A/Accession: T10361
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-279 <AHR>
 A/Cross-references: EMBL:U75930; NID:92934903; PIDN:AMC59091.1; PID:g1911338
 C/Superfamily: proline-rich protein 3

Query Match 5.7%; Score 293; DB 2; Length 279;
 Best Local Similarity 55.8%; Pred. No. 1.e-06;
 Matches 48; Conservative 26; Mismatches 12; Indels 0; Gaps 0;

QY 783 APGSGSPSVSPASPSLSPSPSSSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 842
 DB 78 SPPTSP 137
 QY 843 SP 868
 DB 138 SPPTSP 163

RESULT 14

550754
 hypothetical protein wpe - Chlamydomonas eugametos
 C/Species: Chlamydomonas eugametos
 C/Date: 14-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 21-Jul-2000
 C/Accession: S50754
 R/Wessener, J.P.; Molendijk, A.J.; van Egmond, P.; Klis, F.M.; Goodenough, U.W.; Haring,
 Plant Mol. Biol. 26, 947-960, 1994
 A/Title: Domain conservation in several volvocalean cell wall proteins.
 A/Reference number: S50754; PMID:95093034; PMID:8000007
 A/Accession: S50754
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-351 <MOR>
 A/Cross-references: EMBL:L29028; NID:9530877; PIDN:AAB53954.1; PID:9530878

Query Match 5.7%; Score 291; DB 2; Length 351;
 Best Local Similarity 64.4%; Pred. No. 1.e-06;
 Matches 58; Conservative 10; Mismatches 18; Indels 4; Gaps 2;

QY 783 APGSGSPSVSPASPSLSPSPSSSPSPSP--SPGSPGSPSPSPSPSPSPSPSPSP 840
 DB 175 SP 234
 QY 841 SASP--SP 868
 DB 235 VASPOQSP 264

RESULT 15
T34434
hypothetical protein K06A9.1a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C:Accession: T34434
R:Geisels, C.; Gattung, S.
Submitted to the EMBL Data Library, December 1996
A:Description: The sequence of C. elegans cosmid K06A9.
A:Reference number: 221525
A:Accession: T34434
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2232 <GRI>
A:Cross-references: EMBL:U80846; PIDN:AACT0890.1; GSPDB:GN00028; CESP:K06A9.1a
A:Experimental source: strain Bristol N2; clone K06A9
C:Genetics:
A:Gene: CESP:K06A9.1a
A:Map position: X
A:Insertions: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 1088/1; 1367/1; 2039/1; 2049/1; 2075/1
Query Match 5.7%; Score 290.5; DB 2; Length 2232;
Best Local Similarity 21.8%; Pred. No. 1.6e-05;
Matches 239; Conservative 114; Mismatches 428; Indels 315; Gaps 45;
QY 18 SLAATASFAVAALGVLPPIA---ITASPAHAATQPYTMSNVAI---GGGFGVDGIVFNE 72
DB 235 SLGTSSSPLPSSISITSAIPASSASSSPSAASSTTPVLLSSSTIQSSGTPPSSVASSP 294
QY 73 GAGELIVRTIGGMYRDANGRMIFLIDWVGMMNGVGVSTAD-----PINTMK 126
DB 295 STVG-----STSG-----AASSSYATVTIAGSTGTTTPVPGSS 330
QY 127 VMAIVGWYTNMWDPNDAILR---SSDQATWQITP-LPEKLGAMP---GRGWGERLAVD 180
DB 331 --STTGSTPSSASSSSSGTMTSTGSTVTVVPSSSTPSSSTPIASSSSPGSTVVA 388
QY 181 PNNNNILYFGAPSGKGL---WRSTDSGATWSQMTNFPDGVYIANPTDTGYQSDIQGVV 237
DB 389 PGSSSTYGSSSTPSSASSSSSGTMTSTGSGTIVTAPVSSSTFGSSSTPIASSSSSGSTVT 448
QY 238 WVAFDK-----SSSSLGQAS-----KTIIVGVADPNNPVFWSRDQATWQA 278
DB 449 VVSGSSSTYGSSSTPSSASSSAGTASTIGSTGSTATIVPGSS-----SSVGSSTQSA 500
QY 279 VPGAPGTGTPHKGVDPNNHLYITSNTEGPGYDSSGQVWKFSTVSGTWITISPPSTPD 338
DB 501 SPSSPG-----TMSVTVSGP-TGSTVTVVPSSSTSPA-PSSSPNPS 540
QY 339 TANDYFGYSGELIDRQHNTIMVATQISWMPDTIIFRSTDGATWT---RIMDMTSYPMR 395
DB 541 PAS-----TGSTTTTSSGSSIIIVT-----VGSITVSGSTGTSOSTLASSTATPSS 586
QY 396 SLRYVLDISAE-----PWLTFGVQPNP---PVSPKLGWMDRAMAIDPFNS 438
DB 587 SSTVPSSSSPQSSSQSPAPNTGSTTPSQTSQSSPSPNMPSSTPTGSSQSTITTEGSTA 646
QY 439 DRMLYGTGATLYATNDLTKMDSGGQIHIAWYKGLBETAANDLISPPSGAPLISALGDI- 497
DB 647 SSPGSGTSTGTVAVTEVTSQST-----VPSGSSLGTSQSTNSPSSSISPSTGMSLT 700
QY 498 -----GGFTHADVTAVPS-----TIFSPVFTTGTSDVYALNPSIIIVRGS 539
DB 701 SEBPSSSTQSSGAQSTLTTPSPNPQSTSSLESSTGATSSGAGTTWTSPPQSSSVGS 760
QY 540 FDPSSQPNDRHVAFTSDGAKWFOGS--EPG-GVTTGGTVAASADGSRFVWAPG----- 590
DB 761 SQGSTSP-----AASSTGEMTSQSTQTPGSSVSTSAAILTSTQOSVSTNSPGSTVTRP 815
QY 591 -----DGGQPVVY-----AVFGNSMAASQGVANAOIRSDRVNPKTFYALSNGTFRS 639
DB 816 STVGSSTSSGSTVTVGSTEASTGSSVASSSPAPSTSQ-----NPNPSTSSGSSMITQS 869

QY 640 TDGVTFOFVAAGLPSSGAVGVMFHAVPGKEGDLMLAASGLYHSTNGSSMSA-----IT 695
DB 870 PYPSSQSTSPVSSSTPS-----PSSPGTTLTSTSPSPSQSTTIGSTGSPGIS 919
QY 696 GYSSAANVGFSGKAPSSSYPAVFVGTIGVTAAYRSDDCGTTWVLINDQHYGNWQQA 755
DB 920 TTSEMTSQSTQTPPSTGTVTQPSIVSDT----- 951
QY 756 ITGDHANLRVYIGTNGRGIVYDYGAPSGSPSVSPASPSLSPSPS-----PSSSP 810
DB 952 -----SSGSTVTVGSTEQ--SSSPISPTSGQNTNPSTSSGSSMSTQTPQSSQ 995
QY 811 SPSP---SPSSSPSSSPSP-----SPSPSPSP-----SPASPSPS 848
DB 996 STSPVSSSTGATSSSGSPGTTLTSLISPPSPSSTIGSQGSTSPVVTISOGSTETPGS 1055
QY 849 SPSPSSSP-----SSSPPTSSSPVSGVAVQYKND 881
DB 1056 TGSITTKPSTVSGSASGSGTATMGSTEASTSGSSSTSPNPSQSTSPSTSGA-----T 1108
QY 882 SAPGDNQIKPGLQVNVNTGSSVDLSTVTVRYWFTPRDGSSTLYNC--DMAAIGCNIRA 939
DB 1109 SSPGSS-----GTLTSLISPPSQSSTIG-----SSQGSTSPVVTSTGDMTSGSGTQIPG 1159
QY 940 SPGS--VNPATPTADT 953
DB 1160 STGSTVTPSTGSGST 1175

Search completed: May 14, 2004, 09:18:40
Job time : 32 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 14, 2004, 09:10:23 ; Search time 18 seconds
(without alignments)

2768.397 Million cell updates/sec

Title: US-09-917-376-1

Perfect score: 5135

Sequence: 1 MDRSENIRLTWRSRLVSL.....RASFGSVNPAFTADTYLQX 957

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	397.5	7.7	1742	1 GUNA_CALSA	P22534 caldocellium
2	340.5	6.6	1331	1 MAMB_CALSA	P22533 caldocellium
3	293	5.7	279	1 Y091_NPVOP	O10341 oryza pseu
4	290	5.6	1039	1 GUNB_CALSA	P10474 c endogluc
5	283.5	5.5	474	1 VTP3_TTVIV	P19275 thermoprote
6	272	5.3	555	1 GPT_CHIRE	O9496 chlamydomon
7	261.5	5.1	1853	1 CIPA_CLOTM	O06851 clostridium
8	250	4.9	268	1 NO20_MEDTR	P93329 medicago tr
9	245	4.8	772	1 CIPB_CLOTM	O01866 clostridium
10	245	4.8	879	1 GUNI_CLOTM	O02934 clostridium
11	244	4.8	1664	1 SLPI_CLOTM	O06852 clostridium
12	235.5	4.6	2124	1 PGCA_PAT	P07897 rattus norv
13	222.5	4.3	360	1 VTPX_TTVI	P19274 thermoprote
14	222.5	4.3	5703	1 MUSB_HUMAN	O9484 homo sapien
15	221	4.3	2132	1 PGCA_MOUSE	P10282 mus musculu
16	218.5	4.3	3164	1 TBGU_HSV11	P10220 herpes simp
17	216	4.2	449	1 APG_BRANA	P40603 brassica na
18	214	4.2	532	1 SPG7_DICDI	P22688 dictyostell
19	210.5	4.1	545	1 GUN1_ACICD	P54583 acidotherru
20	209	4.1	535	1 SPKC_SYNY3	P74745 synochocyst
21	209	4.1	2415	1 PGCA_HUMAN	P16112 homo sapien
22	208.5	4.1	485	1 SSGP_VOLCA	P21997 volvox carc
23	207.5	4.0	704	1 GP10_DICDI	O06885 dictyostell
24	205.5	4.0	500	1 GUNA_PAELA	P29739 paenibacill
25	204.5	4.0	499	1 GUN2_BACST	P10475 bacillus su
26	204.5	4.0	534	1 APG_ARATH	P40602 arabidopsis
27	204.5	4.0	1970	1 RPB1_HUMAN	P24928 homo sapien
28	204.5	4.0	1970	1 RPB1_MOUSE	P08775 mus musculu
29	204	4.0	444	1 GUNN_ERWCA	O59393 erwina car
30	203.5	4.0	499	1 GUN1_BACST	P07983 bacillus su
31	201	3.9	456	1 GUNA_MICBI	P26414 microbispot
32	199.5	3.9	499	1 GUN3_BACST	P23549 bacillus su
33	199.5	3.9	1367	1 AMYH_YEAST	P08640 saccharomyc

34	199	3.9	1324	1 IRS2_HUMAN	O9482 homo sapien
35	198.5	3.9	505	1 GUNV_ERWCA	O47096 erwina car
36	198	3.9	467	1 RPB1_CRIGR	P1414 cricetus
37	197	3.8	344	1 YBIE_SCHPO	P81719 schizosach
38	197	3.8	2333	1 PGCA_CANFA	O28343 canis fami
39	195.5	3.8	1419	1 ALAI_CANAL	O13366 candida alb
40	194.5	3.8	1852	1 RPB1_CABEL	P16356 caenorhadi
41	194	3.8	1150	1 APMT_PIG	P12021 sus scrofa
42	193.5	3.8	1260	1 ALS1_CANAL	P46590 candida alb
43	192.5	3.7	902	1 RPB1_DICDI	P35084 dictyostell
44	192	3.7	2364	1 PGCA_BOVIN	P13608 bos taurus
45	191.5	3.7	412	1 CSP_FLARA	P02893 plasmodium

ALIGNMENTS

RESULT 1
GUNA_CALSA STANDARD; PRT; 1742 AA.
ID GUNA_CALSA
AC P22534;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Endoglucanase A precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase A) (Cellulase A).
GN CELA.
OS Caldocellum saccharolyticum (Caldicellulosigranulosa).
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosigranulosa.
OX NCBI_TaxID=44001;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95336703; PubMed=7612247;
RA Te O V.S., Saul D.J., Bergquist P.L.;
RT "celA, another gene coding for a multidomain cellulase from the
RT extreme thermophile Caldocellum saccharolyticum";
RL Appl. Microbiol. Biotechnol. 43:291-296(1995).
RN [2]
RP SEQUENCE OF 1516-1742 FROM N.A.
RX MEDLINE=91247819; PubMed=2039230;
RA Leuthi E., Jasmat N.B., Grayling R.A., Love D.R., Bergquist P.L.;
RT "Cloning, sequence analysis, and expression in Escherichia coli of a
RT gene coding for a beta-mannanase from the extremely thermophilic
RT bacterium Caldocellum saccharolyticum";
RL Appl. Environ. Microbiol. 57:694-700(1991).
CC -1- FUNCTION: THE N-TERMINAL DOMAIN OF CELA ENCODES FOR AN
CC ENDOGLUCANASE ACTIVITY ON CARBOXYMETHYLCELLULOSE. THE C-TERMINAL
CC DOMAIN PROBABLY ACT SYNERGISTICALLY TO HYDROLYZE CRYSTALLINE
CC CELLULOSE.
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -1- PTM: The linker region (also termed "hinge") may be a potential
CC site for proteolysis.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY
CC E (FAMILY 9 OF GLYCOSYL HYDROLASES).
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY
CC I (FAMILY 48 OF GLYCOSYL HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: L32742; AAA91086.1; --
CC EMBL: M36063; AAA72860.1; --
CC EMBL: L01257; -- NOT ANNOTATED_CDS.
CC PIR: T17120; T17120.
CC HSP: P26221; 1TF4.
CC InterPro: IPR001956; CBD_3.

DR InterPro; IPR008965; Cellul bind.
 DR InterPro; IPR000556; Glyco_hydro_48.
 DR InterPro; IPR001701; Glyco_hydro_9.
 DR InterPro; IPR008928; Glyco_trans_6hp.
 DR Pfam; PF00942; CSM_3; 3.
 DR Pfam; PF02011; Glyco_hydro_48; 1.
 DR Pfam; PF00759; Glyco_hydro_9; 1.
 DR PRINTS; PR00844; GLHYDRASE48.
 DR ProDom; PD001947; CBD_3; 2.
 DR ProDom; PD011903; Glyco_hydro_48; 1.
 DR PROSITE; PS00592; GLYCOSYL_HYDROL_F9_1; 1.
 DR PROSITE; PS00698; GLYCOSYL_HYDROL_F9_2; 1.
 DR Cellulose degradation; Hydrolase; Glycosidase; signal; Repeat.
 KM SIGNAL 1 23
 FT CHAIN 24 1742 ENDOLUCANASE A.
 FT DOMAIN 24 642 CATALYTIC 1.
 FT DOMAIN 643 700 LINKER ("HINGE") (PRO-THR BOX).
 FT DOMAIN 701 857 CELLULOSE-BINDING (BY SIMILARITY).
 FT DOMAIN 858 903 LINKER ("HINGE") (PRO-THR BOX).
 FT DOMAIN 904 1060 CELLULOSE-BINDING (BY SIMILARITY).
 FT DOMAIN 1061 1112 LINKER ("HINGE") (PRO-THR BOX).
 FT DOMAIN 1113 1742 CATALYTIC 2.
 FT ACT_SITE 396 396 BY SIMILARITY.
 FT ACT_SITE 434 434 BY SIMILARITY.
 FT ACT_SITE 443 443 BY SIMILARITY.
 FT CONFLICT 1545 1545 T -> A (IN REF. 2).
 SQ SEQUENCE 1742 AA; 193696 MW; 3F0699A2123BED07 CRC64;

Query Match 7.7%; Score 397.5; DB 1; Length 1742;
 Best Local Similarity 21.7%; Pred. No. 4.7e-10;
 Matches 189; Conservative 109; Mismatches 269; Indels 303; Gaps 35;

262 PNNPFFMRDGCATQCA-----VPGAPTFPHKGVDPNNHLY-LATNTGSPRYGSSG 316
 48 FN---WVRN---NWRGDSALKDQDNGDLTGWFPDAGDHKFNPKMYSTGLMSWAAY 100
 317 DVWKEFSVSGTWTRISPVSTDTANDYFGYSGLTIDRHPNTIMVATQI-----SW- 368
 101 EYKDAFVSGGLEHI--LNQIEWANDYF-----VKHPKSKVYVYQVGDGSKDAWNG 151
 369 -----PDTIFRSTDGATWTRIMWTSYPNRSIAYLIDISAPMLTFCVQNPVPV 419
 152 PAEVQMERSPKVTQSSPSGA-----VVAETASIALAASIVLKDN 193
 420 PSPKLGMDENAIIDPNSDRMLYGTGATLYATNDLTCKDSSGQIHIAWVGLSTAYN 479
 194 PTKATYLOHA-----KDLYEPAEVTKSDSG-----TYAAN 224
 480 DLISPPSG-----APLISALGDLGGFTADVTAVPSTIPTSPTFTGTSVDY---- 526
 225 GYNSWMSGFYDELMAAWLYLATNDSTYLTAR-----SYQONWPKISGNSIIDYKMAH 279
 527 -----AELNPSITVRAGSPDPSQPNDRHVAFTDG-----GKNW----- 561
 280 CMDVDVHNGAAL---LLAKITDKDTYKQIIESHLDYTTGNGERIKYTPKGLAMLQWGS 336
 562 -----FQSGRPGVTTGGTVAACADG-SRWVMAFGPGGQGVVAVAGG----- 603
 337 LRYATTTAFLLAVYSPWMSGCPGKTKETKRFESQIDVALGSTRS--FWVGGINPPKR 394
 604 -----NSWMASSQVPA-----NAQIRSD----- 621
 395 PHHRTAHSSWMASSQISYHRTLLYALVGGPSSDSTYDDISNYNNVNEVACYNAGFVG 454
 622 -----RVNP---KTFYALSNCTFYRSTDGVTQPVAVAGLPSGAGVGVFAVPG 668
 455 ALAKMYLLYGGNPIIDPFKAIETPTNDEPF-----VEAGINASGNFIEIKAI-- 501
 669 KEGDMLAASSGLYHSTNGSSWSA-----ITGVSAVNVGFGKASAPSSIPAVF 718
 502 -----VNNQSGMPARATNKLKRFYFVDSLILIKAGYS-----FNQL 537
 719 VVGTI---GGVTGAVRSDDCGTTWVLND-----DQHQYGMWGAITGDHNLRR 765

Db 538 TISTNYQAKVSGPYWSSRNIIYILVDFGTGLIPGQDRIKKEVQRIAPQVQ- 536
 QY 766 VYGTNGRGIVYDIDGAPPS-----PSPSVSPASPSLSPSPSPS 807
 Db 597 ---WDSNDSYFSDIKVSGSVVTKRIPLYHEDIKWMBEFGTS-----GVSPRTAS 648
 QY 808 SSP 867
 Db 649 VPTPTPTPTPTA-----TPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 700
 QY 868 P-VSGGVKQYKKNDAAPGNQIKPGLQVNVNCTSSVDLTVVRYVFTDGGSSITLVN 926
 Db 701 PATSGQIKVLYANKETSTNTITRPMVKVNSGSSSIDLSRYTIRVYVYVGERAGSAIS 760
 QY 927 CDWAAGCGNIRASPGSVNPAFTPTATYLD 956
 Db 761 -DWAQIGASVNTKFKVGLSSVSGADYTLR 789

RESULT 2
 NAME CALSA STANDARD; PRT; 1331 AA.
 ID NAME CALSA STANDARD; PRT; 1331 AA.
 AC P22533;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-JUL-1993 (Rel. 26, last sequence update)
 DT 28-FEB-2003 (Rel. 41, last annotation update)
 DE Beta-mannanase/endoglucanase A precursor [includes: Mannan endo-1,4-beta-mannosidase A (EC 3.2.1.78) (Beta-mannanase) (Endo-1,4-mannanase); Endo-1,4-beta-glucanase (EC 3.2.1.4) (Cellulase)].
 GN MANA.
 OS Caldocellum saccharolyticum (Caldicellulosiruptor saccharolyticus).
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
 OC Caldicellulosiruptor.
 OK NCBI_TaxID=44001;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93119139; Pubmed=1476429;
 RA Gibbs M.D., Saul D.J., Luehi E., Bergquist P.L.;
 RT "The beta-mannanase from 'Caldocellum saccharolyticum' is part of a multidomain enzyme.";
 RL Appl. Environ. Microbiol. 58:3864-3867(1992).
 RN [2]
 RP SEQUENCE OF 1-346 FROM N.A.
 RX MEDLINE=91247819; Pubmed=2039230.
 RA Luehi E., Jasmat N.B., Grayling R.A., Love D.R., Bergquist P.L.;
 RT "Cloning, sequence analysis, and expression in Escherichia coli of a gene coding for a beta-mannanase from the extremely thermophilic bacterium 'Caldocellum saccharolyticum'.";
 RL Appl. Environ. Microbiol. 57:694-700(1991).
 CC -FUNCTION: DEGRADATION OF HEMICELLULOSES, THE SECOND MOST ABUNDANT POLYSACCHARIDES IN NATURE. CONTAINS TWO CATALYTIC DOMAINS WITH MANNANASE AND ENDOLUCANASE ACTIVITIES.
 CC CATALYTIC ACTIVITY: Random hydrolysis of 1,4-beta-D-mannosidic linkages in mannan, galactomannan, glucomannan, and galactoglucomannan.
 CC -CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic linkages in cellulose, lichenin and cereal beta-D-glucans.
 CC -MISCELLANEOUS: This enzyme is most active at pH 6 and 80 degrees Celsius.
 CC -SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL HYDROLASES).
 CC -SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CELLULOSE FAMILY J (FAMILY 44 OF GLYCOSYL HYDROLASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.ebi.ac.uk/announcements> or send an email to license@ebi.ac.uk).

DR EMBL; L01257; AAA71887.1; -
 DR EMBL; M36063; AAA72861.1; -
 DR PIR; A48954; A48954.
 DR HSP; 006851; INBC.
 DR InterPro; IPR001956; CBD 3.
 DR InterPro; IPR008965; Cellul bind.
 DR InterPro; IPR01547; Glyco_hydro_5.
 DR Pfam; PF00942; CBM_3; 2.
 DR ProDom; PD001947; CBD 3; 2.
 DR PROSITE; PS00659; GLYCOSYL HYDROL F5; 1.
 KM Hydrolyase; Glycosidase; Cellulose degradation; Signal;
 KM Multifunctional enzyme.
 FT SIGNAL 1 41
 FT CHAIN 1 41
 FT DOMAIN 42 1331
 FT DOMAIN 42 1331
 FT DOMAIN 326 361
 FT DOMAIN 362 518
 FT DOMAIN 519 564
 FT DOMAIN 565 720
 FT DOMAIN 721 780
 FT DOMAIN 781 1331
 FT ACT_SITE 162 162
 FT ACT_SITE 257 257
 FT CONFLICT 338 338
 FT CONFLICT 340 346
 FT SEQUENCE 1331 AA; 146892 MW; 146892 MW; FFBCE51B8D8F080 CRC64;
 Query Match 6.6%; Score 340.5; DB 1; Length 1331;
 Best Local Similarity 22.7%; Pred. No. 9.6e-08;
 Matches 140; Conservative 97; Mismatches 189; Indels 191; Gaps 29;
 QY 426 WMDAMAIIDPNSDMLVTGATVATNDLTWDSGGQIHAPMYKGLBETVNDLISPP 485
 DB 141 WKELKSVLDG-NEDPVIINIGNEPYGNNTYQW-----VNDTKN-- 178
 QY 486 SGAPLISALGDLGFTADVTAVPS--TFTSPVFTGTSVDYAEINPILVRA--GSF 540
 DB 179 ----AIKALRD-AGFKHTIMVDAFPMGQDMSMTWMDNAGSINEMADPLRLVPSIMHYGVY 233
 QY 541 DPSSQPNDRHVAFFSTDG-----GKNWFGQSEPGVTTGGTVAASADGSFVWAPGDPGP 595
 DB 234 NTASKVEEYIKSFVDKGLPLVIGEFHGHHTD-----GDPDER 270
 QY 596 VV-----YANG-FGNSWMAAGCVANNAQIRS--DRVNKTFFYALNNGTFYESTDGG-- 643
 DB 271 AIVRYAQYKIGLESWMSWGNSVYGYLDVNMNDPNNPTPW-----GQWYITNAIGTSS 325
 QY 644 -----VTFQPVAAGLP-----SSGAVGVNFHAPVGRKGLD--- 673
 DB 326 TPTPTSTVTPPT 382
 QY 674 ---WL---AASSGL-----YHSTNGSSWSAITGV--SAVNVGF-----GKSAP 710
 DB 383 IRPWLKVVNSGSSSIDLSRVTIRYWTVDGERAQSAISDMAQIGASNVTFKFKLSSSVS 442
 QY 711 GSSYPAVAVGTIGVTAAYKSDDCGTTVVLINDP---QHGYGN---WQALITGDHANLR 764
 DB 443 GADY--YLEIFGKSAQGLQPKDGTGEIQMRNPKDMSNYNGNMSWISQMSITS----- 494
 QY 765 RYVLTGNGRGIVYGD-----IGGAPSGSPSPSVSPASPSLSPSPSSPSSPSPSPSPSPSP 820
 DB 495 ---YGENEKVATYIDGVLVWGOEBSGA-----TPAPAP 524
 QY 821 SSSP 879
 DB 525 TATPTPTPTVTPPT 576
 QY 880 NDSAPGDNQIRPGLOVAVNTGSSVDLSYTVVYVYTRDGSSTLYNCAIICGNITFA 939
 DB 577 KEINSTTITITIPWLKVVNSGSSSIDLSRVTIRYWTVDGERAQSAIS--DMAQIGASNVTF 635
 QY 940 SFGSVNPTPTPTADTYLQ 956

DB 636 KFYKSSSVSGADTYLH 652
 RESULT 3
 ID Y091 NPVOP STANDARD; PRT; 279 AA.
 AC 010341;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Hypothetical 29.3 kDa protein (ORF92)
 OS Orygia pseudotsugata multicapsid polydactros virus (OpMV)
 OS Orygia pseudotsugata multicapsid polydactros virus (OpMV)
 OS Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
 OC Nucleopolydactrovirus.
 OX NCBI_TaxId=164623;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=7271300; PubMed=9126251;
 RA Andrews C.H., Russell R.R., Funk C.J., Evans J., Harwood S.,
 RA Rohmann G.F.;
 RT "The sequence of the Orygia pseudotsugata multinucleocapsid nuclear
 RT polydactros virus genome."
 RL Virology 229:381-398(1997).
 CC -1- SIMILARITY: TO CORRESPONDING ORF IN ACMPV.
 CC -1- SIMILARITY: Contains 1 chitin-binding type-2 domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; U75930; AAC5091.1; -
 DR PROSITE; PS50940; CHIT_BIND_II; 1.
 KM Hypothetical protein.
 FT DOMAIN 213 273
 FT SEQUENCE 279 AA; 29289 MW; 6FAADAA01009D80 CRC64;
 Query Match 5.7%; Score 293; DB 1; Length 279;
 Best Local Similarity 55.8%; Pred. No. 2e-06;
 Matches 48; Conservative 26; Mismatches 12; Indels 0; Gaps 0;
 QY 783 APBGSPPSVSPASGSLSPSPSSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSA 842
 DB 78 SPPTPSPTLSP 137
 QY 843 SPSPSSP 868
 DB 138 SPPTPSPTLSP 163
 RESULT 4
 ID GUNB CALSA
 AC P10474;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-JUL-1989 (Rel. 11, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Endoglucanase/exoglucanase B precursor [includes: Endoglucanase
 DE (EC 3.2.1.4) (endo-1,4-beta-glucanase) (Cellulase)
 DE (Cellulobiohydroxylase); Exoglucanase (EC 3.2.1.91) (Exocellulobiohydroxylase)]
 DE (1,4-beta-cellulobiohydroxylase)].
 GN CLTB.
 OS Caldectellum saccharolyticum (Caldectellulosiraptor saccharolyticus).
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
 OC Caldectellulosiraptor.
 OX NCBI_TaxId=44001;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=6098398; PubMed=2789517;


```

DT      28-FEB-2003 (Rel. 41, Created)
DT      28-FEB-2003 (Rel. 41, last sequence update)
DT      28-FEB-2003 (Rel. 41, last annotation update)
DE      Vegetative cell wall protein gpl precursor (Hydroxyproline-rich
DE      glycoprotein 1).
GN      GPl.
OS      Chlamydomonas reinhardtii.
OC      Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC      Chlamydomonadaceae; Chlamydomonas.
OX      NCBI_TaxID=3055;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=21159092; PubMed=11258910;
RA      Ferris P.J., Moesner J.P., Wafenschmidt S., Kitz S., Drees J.,
RA      Goodenough U.W.;
RT      "Glycosylated polypurine II rods-with-kinks as a structural motif in
RT      plant hydroxyproline-rich glycoproteins.";
RL      Biochemistry 40:2978-2987(2001).
CC      [2]
RP      PARTIAL PRELIMINARY SEQUENCE FROM N.A.
RX      MEDLINE=91017504; PubMed=1699225;
RA      Adair W.S., Apt K.E.;
RT      "Cell wall regeneration in Chlamydomonas: accumulation of mRNAs
RT      encoding cell wall hydroxyproline-rich glycoproteins.";
RL      Proc. Natl. Acad. Sci. U.S.A. 87:7355-7359(1990).
CC      -I- FUNCTION: Major component of the outer cell wall w6 (crystalline)
CC      layer.
CC      -I- SUBUNIT: Associates with GP2 and GP3.
CC      -I- PTM: N-glycosylated and O-glycosylated.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; AF309494; AAG45420.1; -.
DR      EMBL; M58496; AAA69706.1; ALT_SEQ.
DR      GlycoSuiteDB; Q9PP06; -.
DR      InterPro; IPR003882; Pistil_extensin.
DR      PRINTS; PR01218; PSTLEXTENSIN.
KW      Glycoprotein; Repeat; Signal.
FT      SIGNAL          1           29
FT      CHAIN          30         555   VEGETATIVE CELL WALL PROTEIN GPI.
FT      DOMAIN        40         379   POLY-X PRO.
FT      DOMAIN        259        279   POLY-PRO.
FT      CARBOHYD       399        399   N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT      CARBOHYD       455        455   N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT      CARBOHYD       493        493   N-LINKED (GLCNAC. . . ) (POTENTIAL).
SQ      SEQUENCE       555 AA; 54219 MW; 6A584A90465502F5 CRC64;

Query Match      5.3%; Score 272; DB 1; Length 555;
Best Local Similarity 44.6%; Pred. No. 3.2e-05;
Matches 74; Conservative 15; Mismatches 41; Indels 36; Gaps 8;

QY      PSGPSPEVSPS--ASPLSLSPSPSS--SPSPSPSPSS-----SPSSPSPS--PSPPS 833
Db      PSSPSPSPSPAAPPPTTPPBDSPPSPVPBPAFPBPAPSPAPSPSPSPAAPPPTSPS 346
QY      PSRSPSPASPSPPSSSPSSSPSSSPSPSSPSPPRSSPVSGGVKQVKNNDSAPDNQIKBL 893
Db      PSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPVA--VLWMDALAFOD----- 397
QY      QVVNTGGSSVDLSITYTYRWTFTRDGGSSTLYNMCMAATGC-GNR 938
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      398 ---LNGTS-----TRPSASRMVGEPDIAGTKCKGNLK 427

RESULT 7
CIPA CILOTM
ID CIPA CILOTM STANDARD; PRT; 1853 AA.
```

AC 006851; DT 01-JUN-1994 (Rel. 29, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cellulosomal scaffolding protein A precursor (Cellulosomal
glycoprotein S1/SL) (Cellulose integrating protein A) (Cohesin).
GN CPA.
OS Clostridium thermocellum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1515;
RN [1]
RP SEQUENCE FROM N.A. AND SEQUENCE OF 29-40 AND 42-43.
RC STRAIN=ATCC 27405 / DSM 1237;
RX MEDLINE=93302508; PubMed=8316083;
RA Genzros U.T., Romaniec M.P.M., Kobayashi T., Hunkisson N.S.,
RA Demain A.L.;
RT "Sequencing of a Clostridium thermocellum gene (cpa) encoding the
RT cellulosomal SL-protein reveals an unusual degree of internal
RT homology.";
RL Mol. Microbiol. 8:325-334(1993).
RN [2]
RP SEQUENCE OF 1820-1853 FROM N.A.
RX MEDLINE=93209931; PubMed=8458632;
RA Fujino T., Beguin P., Aubert J.-P.;
RT "Organization of a Clostridium thermocellum gene cluster encoding the
RT cellulosomal scaffolding protein CtpA and a protein possibly involved
RT in attachment of the cellulosome to the cell surface.";
RL J. Bacteriol. 175:1891-1899(1993).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (12.15 ANGSTROMS) OF 184-321.
RX MEDLINE=972838934; PubMed=90831077;
RA Shimon L.J.W., Bayer E.A., Morag E., Lamed R., Yaron S., Shoham Y.,
RA Frolow F.;
RT "A cohesin domain from Clostridium thermocellum: the crystal
RT structure provides new insights into cellulosome assembly.";
RL Structure 5:381-390(1997).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 368-522.
RX MEDLINE=97076134; PubMed=8918451;
RA Torino J., Lamed R., Chirino A.J., Morag E., Bayer E.A., Shoham Y.,
RA Seitz T.A.;
RT "Crystal structure of a bacterial family-III cellulose-binding
RT domain: a general mechanism for attachment to cellulose.";
RL EMBO J. 15:5739-5751(1996).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS) OF 1213-1361.
RX MEDLINE=98022914; PubMed=9402065;
RA Tavares G.A., Beguin P., Alzari F.M.;
RT "The crystal structure of a type I cohesin domain at 1.7-A
RT resolution.";
RL J. Mol. Biol. 273:701-713(1997).
CC -1- FUNCTION: ACTS AS A SCAPFOLDING PROTEIN IN THE CELLULOSOME. IT
CC PROMOTES BINDING OF CELLULOSE TO THE CATALYTIC DOMAINS OF THE
CC CELLULOLYTIC ENZYMES.
CC -1- SUBCELLULAR LOCATION: Cell surface.
CC -1- DOMAIN: THE COHESIN DOMAINS BIND TO THE DOCKERIN DOMAIN BORN BY
CC THE CATALYTIC COMPONENTS OF THE CELLULOSOME.
CC -1- SIMILARITY: Contains 9 cohesin domains.
CC -1- SIMILARITY: Contains 2 dockerin domains.
CC -----
CC CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/annouce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L08665; -; NOT ANNOTATED_CDS.
DR EMBL; X67506; CAA47840.1; -;
DR PIR; S36859; S36859.
DR PDB; 1ANU; 23-JUL-97.

DR PDB; 1A0H; 08-JUL-98.
 DR PDB; INBC; 26-SEP-97.
 DR InterPro; IPR001956; CBD 3.
 DR InterPro; IPR008965; Cellul bind.
 DR InterPro; IPR002102; Cohesin.
 DR InterPro; IPR002105; Dockerin_1.
 DR InterPro; IPR002048; EF-hand.
 DR Pfam; PF00942; CBM_3; 1.
 DR Pfam; PF00963; Cohesin; 9.
 DR Pfam; PF00404; Dockerin_1; 2.
 DR ProDom; PD001947; CBD 3; 1.
 DR PROSITE; PS00018; EF_HAND; UNKNOWN 1.
 DR PROSITE; PS00448; CLOS_CELLULOSE_RPT; 2.
 DR Cellulose degradation; Cell wall; Glycoprotein; Repeat; Signal;
 KW 3D-structure.
 FT SIGNAL 1 28
 FT CHAIN 29 1853
 FT DOMAIN 29 182
 FT DOMAIN 183 322
 FT DOMAIN 323 363
 FT DOMAIN 364 522
 FT DOMAIN 523 559
 FT DOMAIN 560 704
 FT DOMAIN 724 866
 FT DOMAIN 889 1031
 FT DOMAIN 1054 1196
 FT DOMAIN 1219 1361
 FT DOMAIN 1384 1526
 FT DOMAIN 1548 1690
 FT DOMAIN 1791 1814
 FT DOMAIN 1824 1847
 FT CONFLICT 1615 1615
 FT STRAND 185 188
 FT STRAND 190 191
 FT TURN 195 196
 FT STRAND 198 206
 FT TURN 210 211
 FT STRAND 213 221
 FT TURN 224 226
 FT STRAND 227 234
 FT TURN 236 237
 FT TURN 243 246
 FT STRAND 247 252
 FT TURN 253 256
 FT STRAND 257 263
 FT TURN 265 266
 FT TURN 270 271
 FT STRAND 273 273
 FT STRAND 277 286
 FT STRAND 292 304
 FT TURN 305 306
 FT STRAND 309 309
 FT STRAND 313 315
 FT STRAND 317 319
 FT STRAND 369 375
 FT STRAND 381 382
 FT STRAND 385 385
 FT STRAND 387 393
 FT STRAND 399 400
 FT STRAND 401 403
 FT HELIX 404 410
 FT STRAND 418 428
 FT TURN 430 431
 FT STRAND 434 436
 FT STRAND 438 440
 FT HELIX 441 452
 FT TURN 453 454
 FT STRAND 455 463
 FT STRAND 457 468
 FT TURN 470 471
 FT STRAND 473 482
 FT TURN 483 484
 FT STRAND 488 489

FT TURN 491 492
 FT TURN 494 495
 FT STRAND 498 498
 FT STRAND 503 504
 FT STRAND 509 512
 FT TURN 513 514
 FT STRAND 515 518
 FT STRAND 1220 1224
 FT STRAND 1226 1229
 FT TURN 1231 1232
 FT STRAND 1234 1242
 FT TURN 1246 1247
 FT STRAND 1249 1249
 FT STRAND 1251 1257
 FT TURN 1260 1262
 FT STRAND 1263 1270
 FT TURN 1272 1273
 FT HELIX 1279 1282
 FT STRAND 1283 1288
 FT TURN 1289 1292
 FT STRAND 1293 1299
 FT TURN 1301 1302
 FT TURN 1306 1307
 FT STRAND 1309 1309
 FT STRAND 1313 1322
 FT TURN 1324 1325
 FT STRAND 1329 1342
 FT TURN 1344 1345
 FT STRAND 1348 1348
 FT STRAND 1351 1354
 FT STRAND 1356 1360
 SQ SEQUENCE 1853 AA; 196831 MW; 3ABDDC03ABFC5372 CRC64;

Query Match 5.1%; Score 261.5; DB 1; Length 1853;
 Best Local Similarity 23.2%; Pred. No. 0.00033;
 Matches 117; Conservative 74; Mismatches 159; Indels 155; Gaps 20;

QY 508 VPSTIFPSVPTTG---TSVYAEINPSIIVR---AGSFDPSSQPNDRHYA--FSTDGG 558
 DB 55 VPSKGMANCDPVLGYDPNVLEVEVKGSIIKDPSPKSDSAIYPRKNIIVFLFAEDSG 114
 QY 559 KMWFGSGEPGVTTGTVAASADGSRFWAPGDPGQPVYAVGFGNSMAASQGVPAANOI 618
 DB 115 R-----GTYAITDGG-----VFAT-----IVATV 133
 QY 619 RSDRVNPKTFYALSNCFYRS-----TDGV-----TRQPYAAGLPSGAY--- 659
 DB 134 KSAAPAAITL--LEVGAFAANDLVEISTTFVAGGVNLGSSVPTQP--NVPDGVVVEI 188
 QY 660 -----GVMFHAVPKKEGDLMLAASGLYHSTNGSGSWSAITGVSSAVNYFG 706
 DB 189 GKVTGSVGTVEIPVYRGVPSK-----GIANCDFVFRYDPNVLEIIGIDPDIDVDNP 244
 QY 707 KSAAGSSYP---AVFVGTIGVGTGAYRSDDCGTTWVLINDQHQYGNMGQAITGD--H 760
 DB 245 KSFDTAIYPRKTIIVFLFAEDSG--TGAY-----ATKDGVF 279
 QY 761 ANLRVYIGTNGKIYVGDIGCAPSGSPSVSPSASPSUSLSPSPSSSPSPSPSSP 820
 DB 280 AKIRATVKSAPAGYITFDEVGFPAD-----NDLVEQKVSFIDGVNVGATPTKATP 332
 QY 821 SSSP 880
 DB 333 TMTATPTKSAIATPTR-----BSVPNTPTNTANPTVSGNLKVEFYNS 376
 QY 881 DSAFGDNQIKRGQVNVNTGSSVDLSTVTVRYWFTRDGSGSTLVYNCMAAI----- 932
 DB 377 NPSDTNSINPQFVNTMGSSAIDLSKTLRYVTVVQGOQKQTFW--CDHAALISNGSYN 435
 QY 933 -GCCNITASFGSVNPAITPTDTYIQ 956
 DB 436 GITSNVKGFVWKSSSTNNADTYLE 460

ID	_NO20_MEDTR	STANDARD;	PRT;	268 AA.
AC	P93329;			
DT	15-DEC-1998	(Rel. 37, Created)		
DT	15-DEC-1998	(Rel. 37, Last sequence update)		
DT	15-JUL-1999	(Rel. 38, Last annotation update)		
DE	Early nodulin 20 precursor (N-20).			
GN	ENOD20.			
OS	Medicago truncatula (Barrel medic).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
CC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;			
OC	eurosid 1; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.			
OX	NCSI_TaxID=3880;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=cv. Jemalong;			
RX	MEDLINE=98187258; PubMed=9526510;			
RA	Greene E.A., Erard M., Dedieu A., Barker D.G.;			
RT	"MENOP16 and 20 are members of a family of phytoecyanin-related early nodulins";			
RL	Plant Mol. Biol. 36:775-783(1998).			
CC	-I- SIMILARITY: Contains 1 placococyanin-like domain.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (http://www.isb.ch/announce) or send an email to license@isb-sib.ch .			
CC	or send an email to license@isb-sib.ch .			
DR	EMBL; X99467; CAA67830.1; "			
DR	InterPro; IPR008972; Cupredoxin.			
DR	InterPro; IPR003243; P1cyanin_like.			
DR	Pfam; PF02298; Cu_bind_like; 1.			
DR	Prodrom; PD003122; P1cyanin_like; 1.			
KW	Modulation; Signal.			
FT	SIGNAL	1..22	BY SIMILARITY.	
FT	CHAIN	23..268	EARLY NODULIN 20.	
FT	DOMAIN	23..?	PLASTOCYANIN-LIKE.	
FT	DOMAIN	136..145	POLY-PRO.	
SQ	SEQUENCE	268 AA; 28668 MW; 6FA29C5798C75F91 CRC64;		
Query Match:		4.9%; Score 250; DB 1; Length 268;		
Best Local Similarity		40.3%; Pred. NO. 0.00013;		
Matches	71; Conservative	14; Mismatches	53; Indels	38; Gaps
DQ	725 GVTCAYRRDDCGTTVWLINDDOHQGMGAALTGDH----	ANTLRVYIGTGNGIV-YGD	779	
DB	84 GIRGEHVHDHYDONTVVVLKK-----	TGIHFIFSGKRKCRGLKLAVVMVAAP	131	
DQ	780.IGCAPSGSPSPSVSSASP-----	SLSPSPSSSGSPSPSSPSGSSPSGSPSP	828	
DQ	132 VLSSPPPPPSPPTPRSSSTPIPHPPRRSIPSPSPSSPSGSPSPSRSTPIPHPRKS	191		
DQ	829 --SPEPSPSRSPSPASPSSPSGSSPSGSS--	PSGP--SPTPSSSPVSCGVK	874	
DB	192 PASFPSPSLKSPSPSSPSLAPSDPSVASLADSPSDSPSPAPASPSSGSK	247		
RESULT 9				
CIPB_CLOTM				
ID	_CIPB_CLOTM	STANDARD;	PRT;	772 AA.
AC	Q01866;			
DT	01-OCT-1994	(Rel. 30, Created)		
DT	01-OCT-1994	(Rel. 30, Last sequence update)		
DT	28-FEB-2003	(Rel. 41, Last annotation update)		
DE	Cellulosomal scaffolding protein B (cellulosomal glycoprotein SI/SL)			
DE	(Cellulose integrating protein B) (Fragment).			
GN	CIPB.			
OS	Clostridium thermocellum.			

```
CC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
OC Clostridium.  
OX NCBI_TaxID=1515;  
RM  
RP SEQUENCE FROM N.A.  
RX STRAIN=Y.S;  
RZ MEDLINE=9316373; PubMed=140597;  
RA Poole D.M., Morag E., Lamed R., Bayer E.A., Hazlewood G.P.,  
RT "Identification of the cellulose-binding domain of the cellulosome  
substrate S1 from Clostridium thermocellum YS.";  
RL FEMS Microbiol. Lett. 78:181-186(1992).  
CC -I- FUNCTION: ACTS AS A SCAPFOLDING PROTEIN IN THE CELLULOSE. IT  
PROMOTES BINDING OF CELLULOSE TO THE CATALYTIC DOMAINS OF THE  
CELLULOLYTIC ENZYMES PROBABLY THROUGH THE BINDING OF THE NINE  
REPEATED DOMAINS WITH THE DUPLICATED SEGMENT OF 24 AMINO ACIDS  
PRESENT IN CATALYTIC SUBUNITS OF THE CELLULOSE.  
CC -I- SUBCELLULAR LOCATION: Cell surface.  
CC -I- DOMAIN: THE COHESIN DOMAINS BIND TO THE DOCKERIN DOMAIN BORN BY  
THE CATALYTIC COMPONENTS OF THE CELLULOSE.  
CC -I- SIMILARITY: Contains at least 3 cohesin domains.  
CC -I- SIMILARITY: Contains 2 dockerin domains.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation  
at the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See http://www.isb-sib.ch/announce/  
or send an email to license@sib-sib.ch).  
-----  
DR EMBL; X68233; CAAB312.1; .  
DR HSSP; Q06851; IMBC.  
DR InterPro; IPR001956; CBD_3.  
DR InterPro; IPR008965; Cellul bind.  
DR InterPro; IPR002102; Cohesin.  
DR InterPro; IPR002105; Dockerin_1.  
DR InterPro; IPR002048; EF-hand.  
DR Pfam; PF00942; CBM_3; 1.  
DR Pfam; PF00963; Cohesin; 3.  
DR Pfam; PF00404; Dockerin_1; 2.  
DR ProDom; PD001947; CBD_3; 1.  
DR PROSITE; PS00019; EF HAND; UNKNOWN 1.  
DR PROSITE; PS00448; CIOS_CELLULOSE_RPT; 2.  
KW Cellulose degradation; Cell wall; Glycoprotein; Repeat.  
FT NON TER 1  
FT DOMAIN <1> 80 COHESIN 1.  
FT FT 81 93 LINKER (PRO/THR-RICH) .  
FT DOMAIN 94 240 COHESIN 2.  
FT DOMAIN 241 272 LINKER (PRO/THR-RICH) .  
FT DOMAIN 273 439 CELLULOSE-BINDING.  
FT DOMAIN 440 461 LINKER (PRO/THR-RICH) .  
FT DOMAIN 462 607 COHESIN 3.  
FT FT 710 733 DOCKERIN 1.  
FT DOMAIN 743 766 DOCKERIN 2.  
SQ SEQUENCE 772 AA; 82491 MW; BEF06DE5E094FE10 CRC64;  
  
Query Match 4.8%; Score 245; DB 1; Length 772;  
Best Local Similarity 24.4%; Pred. No. 0.00066;  
Matches 95; Conservative 68; Mismatches 116; Indels 110; Gaps 14
```


DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Cell surface glycoprotein 1 precursor (Outer layer protein B) (S-layer
DB protein 1).
GN OLPB.
OS Clostridium thermocellum.
ON Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1515;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCIB 10682;
RX MEDLINE=93209931; PubMed=8459832;
RA Fujino T., Beguin P., Aubert J.-P.;
RT "Organization of a Clostridium thermocellum gene cluster encoding the
RT cellulosomal scaffolding protein Cipa and a protein possibly involved
RT in attachment of the cellulosome to the cell surface.";
RL J. Bacteriol. 175:1891-1899(1993).
CC -!- SUBUNIT: ASSEMBLED INTO MONO-LAYERED CRYSTALLINE ARRAYS.
CC -!- SUBCELLULAR LOCATION: Cell wall.
CC -!- SIMILARITY: Contains 4 S-layer homology (SLH) domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X67506; CAA47841.1; -;
DR PIR; T18262; T18262.
DR InterPro; IPR008965; Cellul_bind.
DR InterPro; IPR001119; SLH.
DR Pfam; PF00395; SLH_3.
DR PROSITE; PS01072; SLH_DOMAIN; 2.
DR Cell wall; S-layer; Signal; Repeat.
KW CELL SIGNAL 1 28 POTENTIAL.
FT CHAIN 29 1664 CELL SURFACE GLYCOPROTEIN 1.
FT DOMAIN 36 763 4 X 156 AA APPROXIMATE REPEATS.
FT REPEAT 36 191 1.
FT REPEAT 207 363 2.
FT REPEAT 409 565 3.
FT REPEAT 607 763 4.
FT DOMAIN 771 1377 APPROXIMATE TANDEM REPEATS OF
FT T-P-S-D-E-P.
FT DOMAIN 1378 1449 GLY/PRO/SER/THR-RICH.
FT DOMAIN 1453 1494 SLH 1 (INCOMPLETE).
FT DOMAIN 1495 1565 SLH 2.
FT DOMAIN 1566 1625 SLH 3.
FT DOMAIN 1626 1646 SLH 4 (INCOMPLETE).
FT SEQUENCE 1664 AA; 178194 MW; 5F396695BA9FE74B CRC64;
SQ
Query Match 4.8%; Score 244; DB 1; Length 1664;
Best Local Similarity 20.8%; Pred. No. 0.0016;
Matches 185; Conservative 96; Mismatches 316; Indels 292; Gaps 44;
QY 100 LLDWVGNWNGYN---GVVSIADPINTNKVWAVGMYNTSWDPNDGAILRSSDQATW 155
DB 174 LFDWDAETITGYEIQDPLIVAEPLKD---ASVALELDKTKVKVGDITAT----- 223
QY 156 QITPLPKLGNPGRGMRGLAVDPNNDNI--LYFGAPSGKGLWRSTDGATWSQMTNF 213
DB 224 -----IKIENKMFAGYQLNIKYDPTMLEAIELETGSAIAKRTWPT--GGTVLQSDNY 275
QY 214 PDVGTIANPTDTTGYQSDIQGVVWVAFDXSSSLGQASKTIFVGVADPNPV-----FWS 269
DB 276 -----GKTTAVANDV--GAGIINFAYNSLTKYRET--GVAEETGIIGKIGFRV 321
QY 270 RDGCAT-----WQAVPGATGPIPHKGVDPDNNVLYIATNTGGPYDSSGDWVWFSV 323
DB 322 LKAGSTAIRFEDTTAMPGAIEGTY---MFDWYG-----ENIKGYSVVQPGIEVAEGE 370
QY 324 TSGTWTTRISVPSPSTDANDYFGYSLGTIDRQHPTIMVATQI--SWPDDTIIFRSTDCGAT 382

Db 371 EPGEPTTEPVP--TETPVD-----PTPTVTEPVPSELDFSYVIMELD--KT 414
QY 383 WTRIWDWTSYP-----NRSRLYVLIDISAEPLWLTGCVQNPVPSPKLGWDEAMAIIDPF 436
DB 415 KVKVGDIIITATIKIENKMFAGYQLNIKYDPTMLEAIELETGSAIAKRTW-----PV 466
QY 437 NSDRMLYGT--GATLYATNDLTKWDGGQIHTAPMVKGL---BETAVNDLISPPSGAPLI 491
DB 467 TGGTVLQSDNYGKTTAVANDV---GAGIINFAYNSLTKYRETGV-----AET 513
QY 492 SALGDLG-----GETHADVTAVPSTIFTSPVPT--TGTSV-DYAEINLPSIIVRAGSF 540
DB 514 GIIGKIGFRLVKGASTAIRFEDTTAMPGAIEGTMYFDMYGENIKGYSVVQPGIEVAEGE- 572
QY 541 DPSSQP-----NDRHVAFTDGGKKNWFGSGEGVTVTGTGVA 578
DB 573 EPTEPVTETPVDPTPTVTEPVPSELDFSYVIMELDKTK-----VKEGDVIA 622
QY 579 SADGSRFVWAFDGPQPVVYAVGFGNSWAASQGVPAQAQIRSDRVNPKTFYALS----- 632
DB 623 TIRVNNIKNLG-----YQIGI-----KYDPKVLFAFNIEGTGP 656
QY 633 -----NGTFVSTP---GGVTFQPVAGLPSGAGVGMF--HAVPKGKGLMLAAS 678
DB 657 IDEGTWPAVGGTILKNRDYLPVTGVAINNVSIGILNFAAYVYVDDYREBKSGED----- 710
QY 679 SGLYHSTNGSSWSAITGVSSAVNVGFG-----KSAPGSSYPVAVFVVGTTG 724
DB 711 -----TGIIG--NIGFRLVKAEDTTIRFELESMPGS----- 740
QY 725 GVTGAYRSDDCGTTWVLINDQHOYGNWQATIGDHANLRRVYIGTNGRGIYVGDIGGAP 784
DB 741 -IDGTYMLD-----WYL-----NRISG-----YVVIQAP 764
QY 785 ----SGSPSPSVSPASPSLSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPS 837
DB 765 IKAASDEPIPTDPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDE 824
QY 838 PSPASPS 886
DB 825 PTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPS 872
RESULT 12
PGCA RAT
ID PGCA RAT STANDARD; PRT: 2124 AA.
AC P07897;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Aggrecan core protein precursor (Cartilage-specific proteoglycan core
DE protein) (CSPCP).
GN AGC1 OR AGC.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88087070; PubMed=3693370;
RA Doerge K., Sasaki M., Horigan E., Hassell J.R., Yamada Y.;
RT "Complete primary structure of the rat cartilage proteoglycan core
RT protein deduced from cDNA clones.";
RL J. Biol. Chem. 262:17757-17767(1987).
RN [2]
RP REVISION TO 698.
RA Doerge K., Sasaki M., Horigan E., Hassell J.R., Yamada Y.;
RL J. Biol. Chem. 263:10040-10040(1988).
RN [3]
RP SEQUENCE OF 1856-2124 FROM N.A.
RX MEDLINE=86250698; PubMed=2424893;
RA Doerge K., Fernandez P., Hassell J.R., Sasaki M., Yamada Y.;


```
QY 527 ABLNPSIIIVRAGSFDPSQNDHRHVAESTDGGKWFQSGSRPGVTTGCTVAASADGRFV 586
Db 1477 EE-----GLGSVELSLGSPGETDLGSGTSGMVD-VSGSGSGAIDSSGLISPTPEFS--- 1525
QY 587 WAPCDPQPVVYAVFGNSWAAAGVPEAN-AQIRSDRVNPKTFYALNNGTFYRSTDG--- 642
Db 1526 -----GLPSGVAEVSGEVGETGSLSSGAP-----DGSGL 1557
QY 643 -----GVTFQVPA--AG-----LPSSGAVGVMFHVPKGEGLD----- 673
Db 1558 VSGPPTVSLVDRTLVESITLAPTAQAEAGEGSSILEFSGA-----HSGTPDISGLSGSLD 1613
QY 674 -----WLAAS-----SGLYHSTN--GGSSWSAITGVSAVNVGFKSAPGSSYP 715
Db 1614 QSTWQPGWTEASTERPSSPFSGDFSTTDAAGSISITAPTG-----SGETSGLP 1662
QY 716 AVFVGT--IGVGTGAYRSDCGTTWVLINDQHOYGNWQAITGDHANLRVVGITNGR 773
Db 1663 EVTLITSELVEGVTPTVSELG-----HGPSMT-----YTPRLF-EASGE 1702
QY 774 GIVYDGG-----APSGSPSPSVSPASPSLSPSPSSSPSPSPSPSSSP 820
Db 1703 ASAGDLGGPVTIPPGSGVSAPEGSDFSAYPEAGVGVSAPEASSQLSEFPDLHGIT 1762
QY 821 SSS-----PSPSPSPSPSP-----SRSPSPSPSPSPSPSPSPSPSPSPSPSP 869
Db 1763 SASRETLEMTTPTGTEVSSNFWTFOETREGSAAPEVSGSSSTSDIDAGTSGVPPATPM 1822
QY 870 SGGVKVQYKNDAPGDNQIKPGLOVNTGSSSVDLSTVT--VRYWTRDGGSSTLVNC 927
Db 1823 TSGDRTEISGEWD-----HTSEVNVTVSTVPESRWAQSTQHTETL----- 1865
QY 928 DWAAIGCGNIRASFGSNVPA-----TPTADT 953
Db 1866 -----QEIGSPNPSYSGEETQTAET 1885

RESULT 13
VTPX TTVI
ID VTPX TTVI STANDARD; PRT; 360 AA.
AC P19274;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-FEB-1991 (Rel. 17, Last annotation update)
DE Viral protein TPX.
OS Thermoproteus tenax virus 1 (strain KRAL) (TTV1).
OC Viruses; dsDNA viruses, no RNA stage; Lipothrixviridae;
OC Lipothrixvirus.
OX NCBI_TaxID=10480;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90174928; PubMed=2308830;
RA Neumann H., Zillig W.;
RT "The TTV1-encoded viral protein TPX: primary structure of the gene
and the protein.";
RL Nucleic Acids Res. 18:195-195(1990).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC -----
CC EMBL; X14855; CAA33002.1; -.
CC PIR; S12850; S12850.
KW Repeat.
FT DOMAIN 278 353 3 THR-PRO REPEATS REGIONS AND TWO NEAR
FT REPEAT 270 291 IDENTICAL REPEATS.
FT REPEAT 292 301 THR-PRO (N).
FT REPEAT 302 322 THR-PRO (N).
```

```
FT REPEAT 323 332 THR-PRO (N).
FT REPEAT 333 353
SQ SEQUENCE 360 AA; 38214 MW; EFS4PBCB753EDICA CRC64;

Query Match
Best Local Similarity 4.3%; Score 222.5; DB 1; Length 360;
Matches 83; Conservative 55; Mismatches 94; Indels 137; Gaps 17;

QY 581 DGSRFVW-----APGDGPQPVVYAVFGNSWAAAGVPEAN-AQIRSDRVNPKTFYALSNGT 635
Db 48 DG---VWVVKLRSLSPGQSITI-----TASSGTP-----NIDPTIALYNNGS 88
QY 636 FYRSTD-----GGVTFQVPAAG-----LPSSGAVGVMFHVPKGEK 671
Db 89 SYSNLTIVGSPVTIVQDFGCVSAISAYASGDFLVAASPTGFTPSSRLLV----- 139
QY 672 DLWL-----AASGLYHSTNGSSWSAITG--VSSAVNVGFKSAPGSSYPVAVFVGT 722
Db 140 DRWATPTSLDAVGLRLYADTN---DWFGVVRKYINGAQNVSIEQKISG----- 185
QY 723 IGVGTGAYRSDCGTTWVLIND-QHOYGNWQA-----AITGDHANLRVVGITNGRGI 775
Db 186 -----TYSVNVNEIDISQFAAFTDPLVWYLSINGSTANV-KVYKQGNIGT 229
QY 776 VYDGGAPSGSPS-----PSVSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 815
Db 230 VSGNVSTTTPYGNPMAGYGTVDKHYANFVLVPEPDQVTVPISSPSPSTPTPTPTPTPT 289
QY 816 PSSSPS---SPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 861
Db 290 PTYDITYVVFVOTSPSTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 349
QY 862 PTPSSSPVS 870
Db 350 PTPSTTSS 358

RESULT 14
MUSB HUMAN
ID MUSB HUMAN STANDARD; PRT; 5703 AA.
AC Q9HC84; Q00447; Q00573; Q14985; Q15494; Q95291; Q95451; Q14881;
AC Q99552; Q9UE28;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Mucin 5B precursor (Mucin 5 subtype B, tracheobronchial) (High
DE molecular weight salivary mucin MGI) (Sublingual gland mucin).
GN MUC5B OR MUC5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE OF 1-1594 FROM N.A.
RA Chen Y., Di Y.P., Wu R.;
RT "Molecular cloning of the amino-terminal and 5'-flanking region of the
human MUC5B mucin gene.";
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
CC [2]
CC SEQUENCE OF 1-1325 FROM N.A.
RX MEDLINE=99009274; PubMed=9790959;
RA Offner G.D., Nunes D.P., Keates A.C., Afdhal N.H., Troxler R.F.;
RT "The amino-terminal sequence of MUC5B contains conserved
RT multifunctional D domains: implications for tissue-specific mucin
RT functions.";
RL Biochem. Biophys. Res. Commun. 251:350-355(1998).
CC [3]
CC SEQUENCE OF 40-1324 FROM N.A.
RX MEDLINE=99023932; PubMed=9604771;
RA Desseyn J.-L., Buisson M.P., Porchet N., Aubert J.-P., Laine A.;
RT "Genomic organization of the human mucin gene MUC5B: cDNA and genomic
RT sequences upstream of the large central exon.";
RL J. Biol. Chem. 273:30157-30164(1998).
```

[4] SEQUENCE OF 1326-4895 FROM N.A.
RN TISSUE=Placenta;
RX MEDLINE=97166151; PubMed-9013550;
RA Desseyn J.-L., Guyonnet-Duperat V., Porchet N., Aubert J.-P.,
LA Laine A.;
RT "Human mucin gene MUC5B, the 10.7 kb large central exon encodes
RT various alternate subdomains resulting in a super-repeat. Structural
RT evidence for a 11p15.5 gene family.";
RL J. Biol. Chem. 272:3168-3178 (1997).
[5] SEQUENCE OF 4057-4480 FROM N.A.
RN TISSUE=Salivary gland;
RX MEDLINE=97292540; PubMed-9147051;
RA Nielsen P.A., Bennett E.P., Wandall H.H., Therkildsen M.H.,
LA Hannibal J., Clausen H.;
RT "Identification of a major human high molecular weight salivary mucin
RT (MG1) as tracheobronchial mucin MUC5B.";
RL Glycobiology 7:413-419 (1997).
[6] SEQUENCE OF 4721-5703 FROM N.A.
RN TISSUE=Gall bladder;
RX MEDLINE=97293229; PubMed-9164870;
RA Keates A.C., Nunes D.P., Afzal N.H., Troxler R.F., Offner G.D.;
RT "Molecular cloning of a major human gall bladder mucin: complete C-
RT terminal sequence and genomic organization of MUC5B.";
RL Biochem. J. 324:295-303 (1997).
[7] SEQUENCE OF 4809-5687 FROM N.A.
RN TISSUE=Sublingual gland;
RX MEDLINE=96125355; PubMed-8554565;
RA Troxler R.F., Offner G.D., Zhang F., Iontcheva I., Oppenheim F.G.;
RT "Molecular cloning of a novel high molecular weight mucin (MG1)
RT from human sublingual gland.";
RL Biochem. Biophys. Res. Commun. 217:1112-1119 (1995).
[8] SEQUENCE OF 4859-5703 FROM N.A.
RN TISSUE=Placenta;
RX MEDLINE=97347489; PubMed-9201995;
RA Desseyn J.-L., Aubert J.-P., Porchet N., Laine A.;
RT "Genomic organization of the 3 region of the human MUC5B mucin.";
RL J. Biol. Chem. 272:16873-16883 (1997).
CC -1- FUNCTION: Salivary mucin that is thought to contribute to the
CC lubricating and viscoelastic properties of whole saliva.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: Expressed mainly in bronchus glands and also
CC in submaxillary glands, endocervix, gall bladder, and pancreas.
CC -1- PTM: Highly glycosylated.
CC -1- SIMILARITY: Contains 1 TIL (Trypsin inhibitory-like) domain.
CC -1- SIMILARITY: Contains 3 VWFC domains.
CC -1- SIMILARITY: Contains 4 VWFC domains.
CC -1- SIMILARITY: Contains 1 C-terminal cysteine knot-like (CTCK) domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

EMBL; AF107890; AAC33673.1; -;
EMBL; AF086604; AAC67545.1; -;
EMBL; AJ004862; CAA06167.1; -;
EMBL; Z72496; CAA96577.1; -;
EMBL; X74955; CAA52910.1; -;
EMBL; U63836; AAB61398.1; -;
EMBL; U78554; AAC51344.1; -;
EMBL; U78552; AAC51344.1; JOINED.
EMBL; U78553; AAC51344.1; JOINED.
EMBL; U78551; AAC51343.1; -;
EMBL; U95031; AAB65151.1; -;
EMBL; Y09788; CAA70926.1; -;

Genew; HGNC:7516; MUC5B.
DR MIM; 600770; -;
DR GO; GO:0005515; F:protein binding; IPI.
DR InterPro; IPR006208; Cys_knot.
DR InterPro; IPR006207; Cys_knot_C.
DR InterPro; IPR009041; PMP_inhibitor.
DR InterPro; IPR002919; TIL_Cysrich.
DR InterPro; IPR006552; VC_out.
DR InterPro; IPR001007; VWFC_C.
DR InterPro; IPR001846; VWFC_D.
DR Pfam; PF00007; Cys_knot; 1.
DR Pfam; PF01826; TIL; 1.
DR Pfam; PF00093; vwc; 1.
DR Pfam; PF00094; vwd; 4.
DR SMART; SM00214; VWC; 6.
DR SMART; SM00215; VWC_out; 4.
DR SMART; SM00216; VWD; 4.
DR PROSITE; PS01185; CTCK_1; 1.
DR PROSITE; PS01225; CTCK_2; 1.
DR PROSITE; PS01208; VWFC_1; 2.
DR PROSITE; PS01184; VWFC_2; 2.
DR Glycoprotein; Repeat; Signal; Polymorphism.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 5703 MUCIN 5B.
FT DOMAIN 77 225 VWFD 1.
FT DOMAIN 329 386 TIL.
FT DOMAIN 426 580 VWFD 2.
FT DOMAIN 858 918 VWFC 1.
FT DOMAIN 896 1044 VWFD 3.
FT DOMAIN 1457 1603 THR-RICH.
FT DOMAIN 1609 4873 THR-RICH.
FT DOMAIN 5005 5178 VWFD 4.
FT DOMAIN 5353 5425 VWFC 2.
FT DOMAIN 5462 5528 VWFC 3.
FT DOMAIN 5594 5683 CTCK.
FT DISULFID 5594 5646 BY SIMILARITY.
FT DISULFID 5622 5660 BY SIMILARITY.
FT DISULFID 5626 5676 BY SIMILARITY.
FT DISULFID 5643 5678 BY SIMILARITY.
FT DISULFID 5645 5682 BY SIMILARITY.
FT CARBOHYD 145 145 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 201 201 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 254 254 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 402 402 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 516 516 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 806 806 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 930 930 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1277 1277 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1293 1293 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1557 1557 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1775 1775 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2192 2192 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2721 2721 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3419 3419 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3948 3948 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4745 4745 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4901 4901 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4958 4958 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4965 4965 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4987 4987 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 5037 5037 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 5052 5052 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 5156 5156 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 5427 5427 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 5467 5467 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 5506 5506 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 5507 5507 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 5543 5543 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 5553 5553 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 5604 5604 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 5618 5618 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 5662 5662 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 5137 5137 T -> S (in dbSNP:2672788).

```
FT 34 34 /FTID=VAR_014123.
FT 95 100 G -> E (IN REF. 2).
FT 104 104 FPLCEN -> LPLCK (IN REF. 2).
FT 142 142 S -> C (IN REF. 2).
FT 225 225 E -> K (IN REF. 1).
FT 330 330 R -> S (IN REF. 2).
FT 337 337 PL -> T (IN REF. 2).
FT 356 356 E -> N (IN REF. 2).
FT 362 362 E -> K (IN REF. 2).
FT 369 369 G -> R (IN REF. 2).
FT 374 374 MISSING (IN REF. 2 AND 3).
FT 393 394 D -> N (IN REF. 2).
FT 468 469 RT -> TR (IN REF. 2).
FT 512 512 RK -> GR (IN REF. 2).
FT 585 587 L -> P (IN REF. 2).
FT 601 601 GAA -> AH (IN REF. 3).
FT 628 629 A -> S (IN REF. 3).
FT 633 633 DP -> RS (IN REF. 2).
FT 676 676 F -> L (IN REF. 2).
FT 676 676 A -> P (IN REF. 3).

Query Match 4.3%; Score 222.5; DB 1; Length 5703;
Best Local Similarity 21.7%; Pred. No. 0.051;
Matches 183; Conservative 99; Mismatches 311; Indels 249; Gaps 44;

QY 224 TDTGTGSDIQVWVAFDKSSSLGQASKTIFGVADPNPFWSRDGGATWQAVGAP 283
DB 2949 TSTATPSSTPGTWTILTGTHAATTATT-----GST--AIPSS 2987

QY 284 TGFPHKGVDPVNVHLYIATS-----NTGPGYDGGSGDVMKFSVTSCTWTRISVPSTDT 339
DB 2988 PGTAPPKVLTS-QATTPATSTSKATSSSPRTATLPLVLTSTATKSTATSTFTIPST- 3045

QY 340 ANDYFGYGLTDR-QHPNTIMVATQISWPDIIIFRST--DGAWTRIDWTSYPNRS 396
DB 3046 ----LGTGTGTSQNRPPHPMATNSTIHPSSTPET-THTSTVLTTKATTTRATSMSTPSST 3100

QY 397 --LRYVL-DISAEPLTFTGQPNPPVPSKLGWMDRAMAIDPFNSDRLYGTGATLYATN 453
DB 3101 PGTWTILTTTAAATTAALPHGTSPSTPGTWT-----ILPESTTATVPTGST--ATA 3154

QY 454 DLTKWDSGGQIHAPVMVKGLEETAVNDLI-----SPSGAPLISALDGLGF-THAD 504
DB 3155 SSTRTAG-----TLKVLTSATTPVTSRATPSSTPGTATPALASTATTPTATS 3207

QY 505 VTAVPS-----TIFTSVFTTCTSDVYAEIENPSII 534
DB 3208 VTAIPSSSLGTAWRLSQTTPATMTATPSSTPETVHTSTVLTATT- 3257

QY 535 VRAGSF-DPSSQPNDRHVAFTDGGKWFQGSPPGVTGGTVAASADGSRFVWAPCDPG 593
DB 3258 TRTGSVATPSSTPGTAHT-----KVPTTTTGTGATPSS-----SPGTAL 3298

QY 594 QPVYAVGFGNSWAASQGVAPANAQIRSDRVNPKTFVLSNGTFYRSYDGGVTFQPVAAAG- 652
DB 3299 TPV-----WISTTTPTT-----RGSIVTPSSI-----PGYTHATVLTITTTTATGS 3343

QY 653 --LPSSGAVGVMPHAFVPGKEGDLWLAAASGLYHSTNGSGSWAITSVSSAVNVGFGKSP 710
DB 3344 MATPSSS-----TQTSQTPSLT--TTATTITATGTTN--PSSTP 3380

QY 711 GSS--YPAVFVVGITGVGA--YRSDCGTWTW--LINDDQHOYGN----- 751
DB 3381 GTTPIPVLTATTATPAATSSSTVPSALGTTHTPPVNTATTGRLSPSPSPHTVPTA 3440

QY 752 WQAIATGDHANLRVYVIGTNGRIVGVGIDIGGAPSGSPSPSVSPASPSLSPSPSPSSPS 811
DB 3441 WTSATSG-----ILGT-----THTEPSTGSHTPAATGTTQSTPA 3478

QY 812 -PSPSPSPSPSPSPSP-SPSPSPSPSP-PSAPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 863
DB 3479 LSSPHSPSPSPSPSP-PSPGTTTTPGHTRGTSRTTATATATPSKTRTSTLLPSPSPSPSP 3537
```

```
QY 864 -----PSSPSVSGGVKQVQKNDNSAPG-----DNQIKPGLQ 894
DB 3538 TTGCEPQCAWSEWLDYSYMPGP-SGGDFDYTNIRAAAGGAVCEQPLGLECRAQAQGPV 3596

QY 895 VVNTG---SSVDLSTVTVRWFTFDGSSSLVYNCDDAAATGC--GNIRASFGSVNPATP 949
DB 3597 LRELQGVVCESLDFGLVCRN---REQVGKFKMCFYIRVFCNNGYCHPCSTPSTATP 3653

QY 950 TA 951
DB 3654 SS 3655

RESULT 15
PGCA MOUSE
ID PGCA MOUSE STANDARD; PRT; 2132 AA.
AC Q61282; Q64021;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Aggrecan core protein precursor (Cartilage-specific proteoglycan core protein) (CSPCP).
DE AGC1 OR AGC.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Cartilage;
RX MEDLINE=95104847; PubMed=7806222;
RA Walcz E., Deak F., Erhardt P., Coulter S.N., Fuloop C., Horvath P., Doege K.J., Glat T.T.;
RT "Complete coding sequence, deduced primary structure, chromosomal localization, and structural analysis of murine aggrecan.";
RL Genomics 22:364-371(1994).
RN [2]
SEQUENCE OF 211-326 FROM N.A.
RC STRAIN=129/Sv;
RX MEDLINE=95004579; PubMed=7920633;
RA Watanabe H., Kimata K., Line S., Strong D., Gao L.-Y., Kozak C.A., Yamada Y.;
RT "Mouse cartilage matrix deficiency (cmd) caused by a 7 bp deletion in the aggrecan gene.";
RL Nat. Genet. 7:154-157(1994).
RN [3]
INTERACTION WITH FBLN1.
RX MEDLINE=99329059; PubMed=10400671;
RA Asberg A., Adam S., Kostka G., Timpl R., Heinegaard D.;
RT "Fibulin-1 is a ligand for the C-type lectin domains of aggrecan and versican.";
RL J. Biol. Chem. 274:20444-20449(1999).
CC -I- FUNCTION: This proteoglycan is a major component of extracellular matrix of cartilaginous tissues. A major function of this protein is to resist compression in cartilage. It binds avidly to hyaluronic acid via an amino-terminal globular region. May play a regulatory role in the matrix assembly of the cartilage.
CC -I- SUBUNIT: Interacts with FBLN1.
CC -I- SUBCELLULAR LOCATION: Secreted; extracellular matrix (By similarity).
CC -I- DOMAIN: Two globular domains, G1 and G2, comprise the amino terminus of the proteoglycan, while another globular region, G3, makes up the COOH terminus. G1 contains link domains and thus consists of three disulfide-bonded loop structures designated as the A, B, B' motifs. G2 is similar to G1. The keratan sulfate (KS) and the chondroitin sulfate (CS) attachment domains lie between G2 and G3.
CC -I- PM: Contains mostly chondroitin sulfate, but also keratan sulfate chains, N-linked and O-linked oligosaccharides.
CC -I- DISEASE: Defects in AGC1 are the cause of cartilage matrix deficiency (CMD). CMD is an autosomal recessive syndrome characterized by cleft palate, short limbs, tail and snout. Mutation in strain CMD causes absence of aggrecan by truncation of
```

the protein (mutation in the G1 domain).

-I- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.

-I- SIMILARITY: Contains 4 link domains.

-I- SIMILARITY: Contains 1 EGF-like domain.

-I- SIMILARITY: Contains 1 C-type lectin family domain.

-I- SIMILARITY: Contains 1 Sushi (SCR) domain.

-I- SIMILARITY: Belongs to the aggregran/versican proteoglycan family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/> or send an email to license@ebi.ac.uk).

EMBL; LO7049; AAC37670.1; --

EMBL; S73722; AAB32160.1; --

DR EMBL; S73721; AAB32160.1; JOINED.

DR PIR; A55182; A55182.

DR HSSP; P98066; 1TSG.

DR MGD; MGI:99602; Agcl.

DR InterPro; IPR002353; AntifreezeZell.

DR InterPro; IPR007110; Ig-like.

DR InterPro; IPR003006; Ig_MHC.

DR InterPro; IPR003596; Ig_V.

DR InterPro; IPR001304; Lectin_C.

DR InterPro; IPR000538; Link.

DR InterPro; IPR003324; SGXXSG.

DR InterPro; IPR000436; Sushi_SCR_CCP.

DR Pfam; PF00047; Ig; 1.

DR Pfam; PF00059; lectin_c; 1.

DR Pfam; PF02339; SGXXSG; 60.

DR Pfam; PF00084; sushi; 1.

DR Pfam; PF00193; Xlink; 4.

DR PRINTS; PR00356; ANTIFREEZEII.

DR PRINTS; PR01265; LINKMODULE.

DR ProDom; PD000918; Link; 4.

DR SMART; SM00032; CCP; 1.

DR SMART; SM00034; CLECT; 1.

DR SMART; SM00406; IG; 1.

DR SMART; SM00445; LINK; 4.

DR PROSITE; PS00615; C TYPE LECTIN_1; 1.

DR PROSITE; PS50041; C TYPE LECTIN_2; 1.

DR PROSITE; PS50835; IG LIKE; 1.

DR PROSITE; PS00290; IG_MHC; 1.

DR PROSITE; PS01241; LINK; 4.

KW Glycoprotein; Proteoglycan; Lectin; Signal; Sushi; Repeat;

KW Immunoglobulin domain.

FT SIGNAL 1 19

FT CHAIN 20 2132

FT DOMAIN 34 147

FT DOMAIN 170 247

FT DOMAIN 268 349

FT DOMAIN 504 581

FT DOMAIN 602 683

FT DOMAIN 1918 2044

FT DOMAIN 2048 2106

FT DOMAIN 48 140

FT DOMAIN 152 247

FT DOMAIN 253 349

FT DOMAIN 486 580

FT DOMAIN 587 682

FT DOMAIN 685 803

FT DOMAIN 805 1231

FT DOMAIN 1232 1917

FT DOMAIN 1917 2132

FT DOMAIN 51 133

FT DISULFID 175 246

FT DISULFID 199 220

FT DISULFID 273 348

FT DISULFID 297 318

FT DISULFID 509 580

FT DISULFID 533 554

FT DISULFID 607 682

FT DISULFID 631 652

FT DISULFID 1922 1933

FT DISULFID 1950 2042

FT DISULFID 2018 2034

FT DISULFID 2049 2092

FT DISULFID 2078 2105

FT CARBOHYD 126 126

FT CARBOHYD 239 239

FT CARBOHYD 333 333

FT CARBOHYD 387 387

FT CARBOHYD 611 611

FT CARBOHYD 667 667

FT CARBOHYD 1675 1675

FT SITE 1171 1173

SQ SEQUENCE 2132 AA; 222008 MW; 0B2BCDFC6CDBA163 CRC64;

Query Match 4.3%; Score 221; DB 1; Length 2132;

Best Local Similarity 20.9%; Pred. No. 0.021;

Matches 199; Conservative 102; Mismatches 311; Indels 338; Gaps 46;

QY 110 GYGVVSIADPINTNKVAAVGMVYNSWDNDGAILRSSDOGA-----TWQITPL 160

DB 1105 GYVSGIPSGDGTET-----SASGVEDVSGLPGSGGLETSASGVDELGPSTRDSLETSAS 1160

QY 161 PFKLGNMGRGMGERLAVDPNNNDNLYFGAPSGKGLWRSTDGATWSQMTNPPD----- 215

DB 1161 GVDVTGPGSGRDPETSVSGVGD-----FSGLPKGEGLETSASGA--EDLSGLPSGKEDL 1215

QY 216 -----VGTYIANPT-DTTGYQSDIQGVVWVAFDKSSSLGQASKTIIFGVA 260

DB 1216 VGSASGALDFGKLPPTGLSGQTPEVNGFPSPGSG-----EYSGADISGSPSS---GLP 1266

QY 261 DPNNPVFWSRDGGATQAVPGAPTGFIPHKGVDFPNVHLYIATSNVTGPGYDG----- 313

DB 1267 D-----FSGLPFGF-PTVSLVD--STLVEVITATSELEGRTIGIS 1306

QY 314 SSGDWMKFSVTSGTWTRISPVSTANDYFGYSGLTIDRQHPNTIMVATQISWPDITII 373

DB 1307 GSGEV-----SGL-----PLGELDSSAD--ISGL-----PSGTELSCQASGSPDS-- 1344

QY 374 FRSTDGGATWTRIMDWTSPNRSRLYVLDISAPMLTTFGVQNPVPSPKLGMDAMAI 433

DB 1345 SGETSG-----PFDVSGQPFSGSGVS-----EETSGI 1371

QY 434 DPFNSDRMLYGTGATLYATNDLTKWD--SGGQIHAPWKGL-----EET 476

DB 1372 PEISGQPS--GTPDPTT-ATSGVTELNELSSGQPDVSGDGSGLFGSGQSGITSVSGETS 1428

QY 477 AVNDLISPSPGAPL-----ISALGDLGGFTHADVTAVPSTIFT----- 514

DB 1429 GISLSCQPSGFPVPSPGTATRTPDLASGTISGSGSGITFVDTSFVEVPTTFREEG 1488

QY 515 -----SPVFTTGTSDVYAEALNPSIIVRAGSFDPSQPNDRHVAFSTDGKNWF 562

DB 1489 GSVELSGFPSCGTELSGTSGTVDVSEQSSGAISSGLTSPTE-----F 1532

QY 563 QGSBPG-----GVTTGTVAASA-DGSRFVWAPGDP-----GQPVVY 598

DB 1533 SGLPSGVAEVSCEFSGETSGSLPSGAFDGSGLV--SGFPTVSLVDRTLVESTQAPTAQ 1590

QY 599 AVRGNSW-----AASQGVPAQAQRSDRVNPKTFVALSNGTFYRSTDGGVTPQVAAGL 653

DB 1591 EAGGPGSGLFSGAHSNTP-----DISGELSGSLDLSTLQSGQMETSTE-----T 1636

QY 654 PSSGAVGMFHAVPKCGEDMLAASSGLYHSTNGSSWSAITGVSSAVNVFGKSPGSS 713

DB 1637 PSS-----PYFSGDF-----SSTTDVSGESIAATTGSGES---SG 1668

QY 714 YPAVFWGT--IGGVGTGAYRSDCCGTTWVLIINDQHQYGNWGQAITGDHANLARVIG-- 769

DB 1669 LPEVTNLTSELVEGVTETVSOELG-----HGFSMT-----YISRL 1704

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 14, 2004, 09:13:38 ; Search time 49 Seconds
(without alignments)
6162.264 Million cell updates/sec

Title: US-09-917-376-1
Perfect score: 5135
Sequence: 1 MDRSENIRLTRSRRLVSL.....RASFGSVNPATPTADTYLQX 957

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues
Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : SPTREMBL 25:*
- 1: sp archaea:*
 - 2: sp bacteria:*
 - 3: sp fungi:*
 - 4: sp human:*
 - 5: sp invertebrate:*
 - 6: sp mammal:*
 - 7: sp mhc:*
 - 8: sp organelle:*
 - 9: sp phage:*
 - 10: sp plant:*
 - 11: sp rodent:*
 - 12: sp virus:*
 - 13: sp vertebrate:*
 - 14: sp unclassified:*
 - 15: sp rvirus:*
 - 16: sp bacteria:*
 - 17: sp archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2479.5	48.3	882	16 Q82M04	Q82m04 streptomyc
2	2436	47.4	890	16 Q86727	Q86727 streptomyc
3	2420.5	47.1	996	2 Q9A0H0	Q9a0h0 caldicellul
4	2037	39.7	839	16 Q97KK0	Q97kk0 clostridium
5	1831.5	35.7	857	3 Q8TFP1	Q8tfp1 aspergillus
6	1784	34.7	806	3 Q9P4T8	Q9p4t8 agaricus bi
7	1740	33.9	856	3 Q74170	Q74170 aspergillus
8	1625.5	31.7	838	3 Q7Z9M8	Q7z9m8 trichoderma
9	1442	28.1	739	16 Q82K30	Q82k30 streptomyc
10	1173.5	22.9	751	16 Q8P1M5	Q8p1m5 xanthomonas
11	1161	22.6	707	16 Q9WYE1	Q9wye1 thermocoga
12	1151.5	22.4	751	16 Q8P9U5	Q8p9u5 xanthomonas
13	1125	21.9	812	3 Q8J0D2	Q8j0d2 geotrichum
14	431.5	8.4	1779	2 Q52374	Q52374 caldicellul
15	409	8.0	196	3 Q8J1H7	Q8j1h7 agaricus bi
16	409	8.0	1770	2 Q9X3P5	Q9x3p5 caldicellul

17	402.5	7.8	921	2 Q9L8L8	Q9l8l8 caldicellul
18	398.5	7.8	1711	2 P96311	P96311 anaerocellu
19	387	7.5	2014	16 Q7U3X4	Q7u3x4 synechococc
20	385	7.5	1064	2 Q7X2U2	Q7x2u2 uncultured
21	372.5	7.3	473	10 Q39620	Q39620 chlamydomon
22	350.5	6.8	997	2 Q9Z411	Q9z411 bacillus sp
23	346.5	6.7	901	2 Q44562	Q44562 actinomyc
24	336.5	6.6	741	16 Q82QP2	Q82qf2 streptomyc
25	334	6.5	1751	2 Q9A0G4	Q9a0g4 caldicellul
26	335.5	6.3	930	2 Q9RFX5	Q9rfx5 caldicellul
27	325	6.3	1000	2 Q24820	Q24820 thermophili
28	320.5	6.2	991	10 Q94C44	Q94c44 chlamydomon
29	318.5	6.2	875	2 Q9F2B0	Q9f2b0 thioobacillu
30	316	6.2	1091	2 Q8KKF7	Q8kkf7 paenibacill
31	308.5	6.0	611	16 P74375	P74375 synechocyst
32	301.5	5.9	1749	16 P73032	P73032 synechocyst
33	298.5	5.8	1154	16 Q7UXK7	Q7uxk7 synechococc
34	297	5.8	714	16 Q7U5X6	Q7u5x6 synechococc
35	297	5.8	1915	2 Q9RPL0	Q9rpl0 acetivibrio
36	294	5.7	474	5 Q86I54	Q86i54 dictyosteli
37	292	5.7	616	2 Q7X2N2	Q7x2n2 thermomonos
38	291	5.7	351	10 Q39492	Q39492 chlamydomon
39	290.5	5.7	2232	5 Q8IFX6	Q8ifx6 caenorhabdi
40	288.5	5.6	1032	5 P91365	P91365 caenorhabdi
41	286	5.6	261	2 Q9AQG7	Q9a0g7 caldicellul
42	279.5	5.4	825	5 Q7YZW4	Q7yzw4 caenorhabdi
43	279	5.4	170	2 Q9RFX6	Q9rfx6 caldicellul
44	277	5.4	234	12 Q91GH4	Q91gh4 epiphyas po
45	276	5.4	1426	2 Q9X3P6	Q9x3p6 caldicellul

ALIGNMENTS

RESULT 1

Q82M04	PRELIMINARY;	PRT;	882 AA.
AC	Q82M04;		
DT	01-JUN-2003 (TREMBLrel. 24, Created)		
DT	01-JUN-2003 (TREMBLrel. 24, Last sequence update)		
DT	01-OCT-2003 (TREMBLrel. 25, Last annotation update)		
DE	Putative endo-1,4-beta-glucanase.		
GN	CELA3 OR SAV1856.		
OS	Streptomyces avermitilis.		
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;		
OC	Streptomycineae; Streptomycetaceae; Streptomyces.		
OX	NCBI_TaxID=33903;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;		
RX	MEDLINE=21477403; PubMed=11572948;		
RA	Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,		
RA	Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,		
RA	Kikuchi H., Shiba T., Sakaki Y., Hattori M.;		
RT	"Genome sequence of an industrial microorganism Streptomyces		
RT	avermitilis: deducing the ability of producing secondary		
RT	metabolites."		
RL	Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).		
RL	[2]		
RN	SEQUENCE FROM N.A.		
RP	STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;		
RC	MEDLINE=22608306; PubMed=12692562;		
RX	Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,		
RA	Sakaki Y., Hattori M., Omura S.;		
RT	"Complete genome sequence and comparative analysis of the industrial		
RT	microorganism Streptomyces avermitilis."		
RL	Nat. Biotechnol. 21:526-531(2003).		
DR	EMBL; AP005028; BAC69567.1; --		
DR	GO; GO:0004553; F:hydrolyase activity, hydrolyzing O-glycosyl . . . ; IEA.		
DR	GO; GO:0005975; P:carbohydrate metabolism; IEA.		
DR	InterPro; IPR001919; Bac celase-bind.		
DR	InterPro; IPR008965; Cellul bind.		
DR	InterPro; IPR002860; GH_BNR_		

DR Pfam; PF02012; BNR; 10.
DR Pfam; PF00553; CBM_2; 1.
DR SMART; SM00637; CBD_II; 1.
KW Complete proteome.
SQ SEQUENCE 882 AA; 92152 MW; A3C8E753CB69F13D CRC64;

Query Match 48.3%; Score 2479.5; DB 16; Length 882;
Best Local Similarity 50.1%; Pred. No. 1.7e-109;
Matches 476; Conservative 129; Mismatches 230; Indels 115; Gaps 15;

QY 11 MRSRLVSLAATAAFVAAL--GVLPITAITASPAHAATTPQVTSNVAIGGGFVDGI 68
DB 1 MRTRIF-----TAVLALAGLPAGTPPALAAPTATIAADTYSKNARVDGGFVPGI 55
QY 69 VFNEGAPGILYVTRTDIGGMYRDAANGRWIPLDWDGNNWNGYGVVSIADPINTKNW 128
DB 56 VFNRSKLNAYARTDIGGAYRMAESSKTWPLDSDVMSDGHGTGVVSLASDSDVPNKV 115
QY 129 AAVGMTNSWDPNDGAILRSDGATWQITPLPFKGGNPGRGMRGLAVIDPNDNLY 188
DB 116 AAVGTYTNSWDPNGAVLRSGDRGASQKTDLPFKLGNPNPGRGMRGLAVIDPNSVLY 175
QY 189 FGAPSGKGLWRSDSGATWQMTNFPDVGTYIANPTDTTCYQSDIQGVVWVAFDKSSSL 248
DB 176 LGAPSGKGLWRSDSGASQVTDFFNVGYVQDADTSGYASDNOQGVVWVTFDESGSP 235
QY 249 GQASKTIFGVADPNPNPFFVRSRGGATWQAVGAPTGFIPKHGVPFVNVHLYIATNTG 308
DB 236 GSTRIYVGVADKNSVYSTDAGATWSLGAQPTGHLAKGVLDAAAGCLYLAYSXDKG 295
QY 309 GPYDGGSDVWKFSTVGTWTRISPVPESTDANDFYSGGLTIDRQHPNTIMVATQISW 368
DB 296 GPYDGGKGLWRVTTKTGTWNTISPAVAEDT---YGFSGLTIDRQHPGTWATAYSSW 352
QY 369 PDTIIFRSTGGATWTRIMWTSPYRSLRYVLDISAEPLWTFGVQPNPVPKLGMD 428
DB 353 PDTQLFRSTDSGGTWTAKWDYTSYPKSNRFTMDVSSPLWTGANPAPPEQPKLGWMT 412
QY 429 EAMAIDPFNSDRMLYGTATLYATNDLTWKDSCGQIHIAPMWGLBETAVNDLISPPS-G 487
DB 413 ESLIEDPFDARSMMYGTCAVYGTDLNLTWDSQSFQTIKPMARGLBETAVNDLASPPSG 472
QY 488 APLISALDGLGFTHADVTAVPSTIFTSVPTFTGTSVDYAEALNPSIIVRAGSFDPPSQPN 547
DB 473 AQLFSAIDGLGFRHTDLTTPVSLMYSNPFTTSTSLDYAETDPGTVVRYGNLD--SGP- 529
QY 548 DRHVAISTDGGKWFQSGPEGVTGTGTAASADGSRFVWAPDGPQPVYVAGFGNSWA 607
DB 530 --HVAFTDNGANWFAGADSPGSGGTVAASDGSFRFVWSPAGTG--VQYTTGFGTWS 585
QY 608 ASOGVPANAQIRSDRVNPKTFYALSNGTFYRSTGDTGVTQPVAA--GLPSSGAVGVMFHAV 666
DB 586 ASAGLPAGALVESDRVDPKTFYFKSGRFVYSSDGGATFTASATGLPDSGS--VRPKAL 643
QY 667 PGKEGLWLAAASS-----GLYHSTNGSSWSAITGVSSAVNVFGFKSAPGSSYPAVFVVG 721
DB 644 PGTKGDIWLAGASDAGYGLWHSTDGGAAFTKLATVDQATDIGFKAATGASQYTLTSA 703
QY 722 TIGGVTVAYRSDCGTTWVLINDHQYGNWQAITGDHANLARVYIGTNGRGIVYDGI 781
DB 704 KIGGVGRIFRSTDKGASWTRVNDDAHQWGTGAITGDPRVYGVVYSTNGRGIVYDGA 763
QY 782 GAPSGSPSPVSPASPSLSLSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPS 841
DB 764 GSSDGG----- 769
QY 842 ASPSPSSSP 901
DB 770 -----GTEAPT-----CACTVYTRITQWGGFO--ADVQLANTGST 805
QY 902 SVDLSTVTVRYWTRDGGSSSTLVNCDWAAIGCG-----NIRASFGS 943
DB 806 AWDGWSLG---WSFGDQOEVTQLWNASYAQAGSGVTAANLAWNGRVAAGS 852

RESULT 2
ID 086727 PRELIMINARY; PRT; 890 AA.
AC 086727;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative secreted cellulase.
GN SCO6545 OR SC5C7.30C.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Parkhill J., Barrell B.G., Ralmandream M.A.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
Huang C.-H., Kieser T., Lark L., Murphy L., Oliver K., O'Neill S.,
Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
Seeger K., Saunders D., Sharp S., Squares S., Squares S., Taylor K.,
Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL; AL939128; CAA20642.1; --
DR PIR; T35237; T35237.
DR HSP; P07986; 1EXG.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001919; Bac celose-bind.
DR InterPro; IPR008965; Cellul_bind.
DR InterPro; IPR002860; GH_BNR.
DR Pfam; PF02012; BNR; 10.
DR Pfam; PF00553; CBM_2; 1.
DR SMART; SM00637; CBD_II; 1.
KW Complete proteome.
SQ SEQUENCE 890 AA; 93252 MW; D2C58695A4B56E84 CRC64;

Query Match 47.4%; Score 2436; DB 16; Length 890;
Best Local Similarity 48.6%; Pred. No. 1.9e-107;
Matches 475; Conservative 129; Mismatches 246; Indels 126; Gaps 19;

QY 11 MRSRLVSLAATAAFVAALGVLPITAITASPA--HAATTPQVTSNVAIGGGFVDGI 68
DB 1 MRTRITVLLALAAAGLLA---GSPFAASAAEPAPRAAADAADTYTKNARIDGGFVPGI 57
QY 69 VFNEGAPGILYVTRTDIGGMYRDAANGRWIPLDWDGNNWNGYGVVSIADPINTKNW 128

Db	58	VFNRTKDLA	YARTDIG	GAYRWQ	BESH	TWTPL	LDDH	VGDW	DWGH	TGV	VALAS	DAVDP	DRVY	117																																										
Qy	129	AAVGMYTNS	WPNDC	GAIR	SLRR	SOGA	TWIT	PLP	FKL	GGN	MPRG	MGER	LAVDP	PNNDN	188																																									
Db	118	AAVGTIYND	WDP	TNGA	VR	ADR	SGAS	WE	KADL	PLP	FKL	GGN	MPRG	MGER	LAVDP	PNNDN	177																																							
Qy	189	FGAPSGKGL	MRS	TDSG	ATWS	QMTN	FPD	VTG	TYI	AN	PTDT	TTGY	QSDI	QGVV	VVA	AFDKSS	-SS	247																																						
Db	178	LGAPSGHGL	MRS	T	DAGV	TWSE	VTAF	PNF	NGYA	QAD	PND	TSG	YASD	NOG	ITW	TFDE	STGG	237																																						
Qy	248	LQOASKTI	FV	GAD	PNN	PV	FW	SR	DGGA	TQ	AV	PG	APT	GPI	PHK	GV	PD	PVNH	VLYI	ANS	WT	307																																		
Db	238	AGTATRT	LY	V	GAD	KEN	AV	YR	STR	DAG	AT	WER	L	AG	Q	TGY	L	AH	K	GV	L	DA	EN	G	Y	L	Y	L	A	S	D	T	297																							
Qy	308	GGPYD	GSSG	DV	WK	SV	TSG	WTR	I	SP	VS	ST	D	AND	Y	FG	Y	S	G	L	T	I	D	R	O	H	E	N	T	I	W	A	T	Q	1	367																				
Db	298	GGPYD	G	G	K	R	U	Y	A	T	A	T	A	T	A	T	A	T	A	T	A	T	A	T	A	T	A	T	A	T	A	T	A	T	354																					
Qy	368	WPD	T	I	F	R	S	T	D	G	G	A	T	W	R	I	D	W	T	S	P	N	R	S	L	R	V	Y	L	D	I	S	A	E	P	W	L	T	F	G	V	Q	N	P	P	V	P	S	K	L	G	M	427			
Db	355	WPD	T	Q	I	F	R	S	T	D	G	A	T	W	S	A	N	S	Y	T	S	P	D	R	E	N	Y	T	M	D	V	S	S	P	W	L	T	G	A	N	P	A	P	P	P	Q	T	K	L	G	M	414				
Qy	428	DEA	M	A	I	D	P	F	N	S	R	D	L	Y	G	T	A	L	I	N	D	L	T	K	W	-	D	S	G	Q	I	H	I	A	M	V	K	G	L	E	E	T	A	N	D	L	I	S	P	S	486					
Db	415	TEA	L	E	I	D	P	F	D	S	R	M	Y	G	T	A	I	N	D	L	T	K	W	-	D	S	G	Q	I	H	I	A	M	V	K	G	L	E	E	T	A	N	D	L	I	S	P	S	474							
Qy	487	GAP	L	I	S	A	L	G	D	I	G	G	F	H	A	D	V	A	P	S	T	I	F	T	S	P	V	T	T	G	T	S	V	D	Y	A	E	L	N	P	S	I	I	V	R	A	G	S	P	D	S	Q	546			
Db	475	GAP	L	L	S	A	L	G	D	V	G	G	F	R	H	T	S	L	T	E	V	P	S	M	Y	T	S	P	N	F	T	S	T	S	L	D	F	A	E	T	K	D	P	V	V	R	A	G	N	L	D	-	S	G	P	532
Qy	547	NDR	H	V	A	S	T	D	G	K	W	F	Q	S	E	P	G	V	T	T	G	T	A	A	S	A	D	G	S	R	P	W	A	P	G	D	Q	P	V	Y	A	V	A	G	E	N	S	W	606							
Db	533	---	H	I	A	F	S	T	D	N	G	A	N	F	G	T	D	S	G	V	G	G	T	A	A	G	D	G	S	R	P	W	S	P	-	-	E	G	A	G	V	Q	T	T	G	F	G	T	S	W	587					
Qy	607	AAS	Q	C	V	P	A	N	A	O	I	R	S	D	R	N	P	K	T	F	Y	A	L	S	N	G	T	E	R	S	T	D	G	V	T	F	Q	P	V	A	A	-	G	L	P	S	S	G	A	V	G	M	F	H	A	665
Db	588	QAS	T	G	L	P	A	G	A	I	V	E	S	D	R	N	P	A	T	P	I	F	G	K</																																

RESULT 3

РЕЗУЛ
09АОН0

ID Q9AQH0

AC Q9AQH0;

DT 01-JUN-2001 (TReMBLrel. 17, Created)

DT 01-JUN-2001 (TREMBLrel. 17, Last seq

DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)

DE Glycosyl hydrolase 5 (Fragment):

OS Caldicellulosiruptor sp. Tok7B.1.

OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae
OC	Caldicellulosiraptor.
OX	NCBI_TaxID=80339;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=Tok7B.1;
RX	MEDLINE=20171169; PubMed=10706665;
RA	Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,
RA	Bergquist P.L.;
RT	"Multidomain and multifunctional glycosyl hydrolases from the extremophile
RT	thermophile Caldicellulosiraptor isolate Tok7B.1.";
RL	Curr. Microbiol. 40:333-340(2000).
RL	EMBL; AF078038; AAK06388.1; -.
DR	HSSP; Q06851; INEC.
DR	GO; GO:0016787; F:hydrolase activity; IEA.
DR	GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR	InterPro; IPR001956; CED_3.
DR	InterPro; IPR008965; Cellul. bind.
DR	InterPro; IPR002860; GH_BNR.
DR	Pfam; PF02012; ENR; 9.
DR	Pfam; PF00942; CBM_3; 1.
DR	ProDom; PD001947; CBD_3; 1.
KW	Hydrolase.
FT	NON TER 996
SO	SEQUENCE 996 AA; 108275 MW; 3C72B6ED22F3C614 CRC64;

RESULT 3

РЕЗУЛ
09АОН0

ID Q9AQH0

AC Q9AQH0;

DT 01-JUN-2001 (TReMBLrel. 17, Created)

DT 01-JUN-2001 (TREMBLrel. 17, Last seq

DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)

DE Glycosyl hydrolase 5 (Fragment):

OS Caldicellulosiruptor sp. Tok7B.1.

```
Qy 666 VPKEGDLMLAASSGLYHSTNGSSWSAITGVSSAVNVGFGKAPGSSYPVAFVVGTTIG 725
Db 646 VPIEGDMLVGNNGMWSHSDGYSFVKISGVEDAASIGFKPAEGETYPALITYAKING 705
Qy 726 VTGAYSDDCGTTWLINDDOHQYGNWQAITGDHANLRVYIGTNGRGIVYDGGAPS 785
Db 706 VRGIFSDDCDKTWRINDDKHOFGCANADITGDPVYGRVFAVATNGLGKWGEIA--Y 762
Qy 786 GSPSPVSPASPSLSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 845
Db 763 SNISPAATPSTPSTPSTPSTPSTPSTPSTPSTPSTPSTPSTPSTPSTPSTPSTP 822
Qy 846 PSSSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 904
Db 823 MTPTPTP-----TPPTPTPTGTCGLKLVLYKNNETSASAGSIRWFVKVNGSSVD 876
Qy 905 LSVTVRYWFTRGGSSTLYNCWAAAGCNGTRASFSGSNVATPTADTYLQ 956
Db 877 LSRVKIRYWTVDGKQPQSAV-CDWAQIGASNTVFNFVKLSSGVSGADYYLE 927

RESULT 4
Q97KK0 PRELIMINARY; PRT; 839 AA.
AC Q97KK0;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Probably secreted sialidase, several ASP-boxes and dockerin
DE domain.
GN CAC0919.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hatti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucet-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum."
RT J. Bacteriol. 183:4823-4838 (2001).
DR EMBL; AE007608; AAK78895.1; -.
DR PIR; D97013; D97013.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0000272; P:polysaccharide catabolism; IEA.
DR InterPro; IPR002105; Dockerin_1.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR002860; GH_BNR.
DR Pfam; PF00212; BNR; 9.
DR Pfam; PF00404; Dockerin_1; 2.
DR PROSITE; PS00018; EF_HAND; 2.
KW Complete proteome.
SQ SEQUENCE 839 AA; 90824 MW; 6C99A041CA9CF984 CRC64;

Query Match 39.7%; Score 2037; DB 16; Length 839;
Best Local Similarity 49.5%; Pred. No. 1.3e-88;
Matches 392; Conservative 118; Mismatches 252; Indels 30; Gaps 16;

Qy 13 SRLVSLAA-----TASFVAALGVLPIAITASPAHAAT-TQPYTWSNVAIIGGGFVD 66
Db 2 NKRIVSMVAGLSIIFTGF-----VTHISAANKAAASVQSQYKWDNAKIGAGGYVP 54
Qy 67 GIYPNEGAPILYVRTDIGMYRWDAANGRWIPLDVGWNNWYGVGVIAADPIYTNK 126
Db 55 AVLFNKTEFDLIYARTDMGAYRWKANKWIPITD--GFSWTMLGCESIATDPIDTNR 112
Qy 127 VAAAVGMYTNSWPNDAIGALLRSSDQATQITLTPFLKLGNNMPGRNGERLAVDPNDNI 186
```

```
Db 113 VYIAGLYTNDQDENAVILSSQDKNGTWKRYQLPFPKVGNNPGRNMGRLQLDPNDK 172
Qy 187 LYFCAPSGKGLWRSTDGATWSQMTNFPDVGTYIANPTDTTGYQSDIQGVVWVAFDKSS 246
Db 173 LYLARSNGGLKMSDEYQOTWSKVDNFPDGTGYVDQPQNE--YTADKVGVMWETFDPSG 230
Qy 247 SLGQASKTIFFGVAD-PNNPVFMSRDGGATWQAVPGATGTFPHKGVFDPVNVHLVIAT 305
Db 231 TKGSPTQTMVGAADKTGNNIVTNDGGKTSVSAVAGQPKGYLPHHGIL-ASDGLMYSIS 289
Qy 306 NTGSPYDSSGDVWKFSTGTRISVPSTDTTANDYFGYSLGTLIDRQHPNTIMATQI 365
Db 290 NTCGPYDSDQVWKYNTKTEWNTITPPAVGDTKS--GFGGI SVDAQNPNNVVVATLN 346
Qy 366 SSWPDTTIFRSTGCGATWTRTWDMTSYPNRSLRYVLDISAEPLWTFG-VQPPPPVPSPKL 424
Db 347 RWFDEBELYRSTDAGTKWKPWDWNGYFNRTLCYNLDYSAQPLDWGKTGVTFPDPLVKL 406
Qy 425 GWMDEAMAIIDPFNSDRMLYGTGATLYATNLDLTKWDSGGQIHIAPMVKGLEETA VNDLI 484
Db 407 GWMGDLLEIDPFNSDRMFPYGTATLYGTDDLTNWDKGNVDISVKANGIEECVNDVVVP 466
Qy 485 PSGAPLISALDGLGFTHADVTAVPSTIFTSPVFTTGTSDYAEINPSSIIVRAGSFDPS 544
Db 467 TKGAQLLSAVGDDCGFYHDDITKVPKMTTFNFSATTSIDYAESVPNFVRVGNVDTSK 526
Qy 545 QPNDRHVAFTSDGGKNWFO-GSEPGGVTTGGTVAASADGSRFPWAPGDPQGPVYAVGFG 603
Db 527 NQDKDCGISYDGGKNWFSAGNISGVYKAGTVAAGADAKTIVWSP-EGANNAAYTDNG 585
Qy 604 NSWAASQGVPANAQIRSDRVNPKTFYALSNGTFYRSTDGGVTF-QPVAAGLPSSGAVGM 662
Db 586 NKWTPCSGLPQGAQVSRDVRNPKFYGLNGKFYISTDAGATFTQSSQTGLPKTKG-GI- 643
Qy 663 PHAVPGKEGDLWLA-ASSGLYHSTNGSSWSAITGVSSAVNVGFGKAPGSSYPVAFVVG 721
Db 644 PKTVIGHEDIIWAGKGLWHSTDSGATFTKVGSDVASDTVGLGKSKTDDGYPALYMDA 703
Qy 722 TIGGVTGAYRSDDCGTTWLVINDDOHQYGNWQAITGDHANLRVYIGTNGRGIVYDGL 781
Db 704 TIDGTAGIFRSDDEGATWVRINDDAHQYGPSDYCITGDPNKYGRVFGVNGRGIVYDGL 763
Qy 782 GAPSGSPSPSPSVS 793
Db 764 GS-QPTPTPSVT 774

RESULT 5
Q8TFP1 PRELIMINARY; PRT; 857 AA.
AC Q8TFP1;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Endoglucanase C.
GN EGLC.
OS Aspergillus niger.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5061;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21914059; PubMed=11916668;
RA Haepser A.A., Dekkers E., van Mil M., van de Vondervoort P.J.I.,
RA de Graaff L.H.;
RT "EglC, a New Endoglucanase from Aspergillus niger with Major Activity
RT towards Xyloglucan."
RL Appl. Environ. Microbiol. 68:1556-1560 (2002).
DR EMBL; AY040839; AAK7727.1; -.
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
```

DR	InterPro; IPR000254; CBD_fungal.	
DR	InterPro; IPR002860; GH_BNR.	
DR	Pfam; PF02012; BNR; 7.	
DR	Pfam; PF00734; CBM_1; 1.	
DR	ProDom; PD001821; CBD_fungal; 1.	
DR	SMART; SM00236; fCBD; 1.	
DR	PROSITE; PS00562; CBD_FUNGAL; 1.	
SQ	SEQUENCE 857 AA; 90459 MW; 402253ECDU919C511 CRC64;	
Query Match 35.7%; Score 1831.5; DB 3; Length 857;		
Best Local Similarity 44.4%; Pred. No. 7.2e-79;		
Matches 382; Conservative 126; Mismatches 306; Indels 47; Gaps 18;		
Qy	45 HAATQYTWSNVAI--GGGFGVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDW 103	
Db	18 HAAASQAYTWKVVVVGCGGGFTPGIVFNPSAKGVAYARTDIGGAYRLN--SDDTTPLMDW 76	
Qy	104 VGNNWNGYGVSTAADPIINTNKVAAVGMVTSNDPNDGAILRSSDQGAQTQIITPLPFK 163	
Db	77 ANNSNHDWGIDATAPVDTDRVYVAVGMTVNDWDNDGSLRSTDOGDTWESTKLPFK 136	
Qy	164 LGGNMPGRGMRGLAVDPNNDNILYFGAPSGKGLWRSTDGATWSQMTNFPDVGTIYANP 223	
Db	137 VGGNMPGRGVRGLAVDPNDNSILYFGARSGNGLWKSTDYGETWSNVTAFKWTGYFQDS 196	
Qy	224 TDTTGYQSDTQGVVWVAFDKSSSLQASKTIYFVGADPNVPFWSDGGAATWQAVPGAP 283	
Db	197 SST--YTSDPVGIAWVTFDSTSGSSGSGFTPRIFVGVVDTCGESVFVEDAGETWTVWSGE 254	
Qy	284 T-GPIPHKGVFDPNVHVLXLTATNTGPGYDGSSGDVWKFSVTCGTWTRISPVPSSTDAND 342	
Db	255 MYGLPHKGLISPEHTLYLISYNGAGPYDGTNGTVHKNTTSQVWTDITSPTSMTDI--- 311	
Qy	343 YFGYSGLTIDRQHPNTIMVATQISWMPDTIIFRSTDGATWTRIDWTSYPNRSLRYLD 402	
Db	312 YYGGLAVDLQVGTWVAALNCWMPDELIWRSTDGSGTWSPIWANGYPSINYYISYD 371	
Qy	403 ISABPWLTFGQVNP-PVPSPKLGWDBEAMAIIDPNSDRMLYGTGATLYATNDLTKWDSG 461	
Db	372 ISNAPLQDDTSTDDBFV--RVGMWVEALAIIDPFDSDHMLYGTGETIYGGHDIQNWDSE 428	
Qy	462 GQIHAPMWKLETAANDLISPPSGAPLISALDGLGFTHADVTAVPSTIFTSPVFTTG 521	
Db	429 HNVTIESLAVGIEEMAVLGLITPPGPAALISAVGDCGFVHTSLUTTAPSQYHYTPTYSST 488	
Qy	522 TSVDYAEINPISIIVRAGSPFPSSQPNDRHVAFTSDGKNWFCQSEPGVTTGGTVAASAD 581	
Db	489 NGIDYAGNKANIVRSGSD----SDPTLALSSFGESWYADYAASSSTATGQVALSAD 543	
Qy	582 GSRFVWAPGPGPVYAVGFGNSWAASQGVPAQAQIRSRVNPKTPIYALNSGTFYFST 641	
Db	544 ADTILIMNSDG----AYRSANSTLSAVSLPSGAVTASDKANNITYFGASGGSFYLSDD 599	
Qy	642 GGVTFOVPAAGLPSSGAVGVWFAVPCGKDWLAASSGLYHSTNGSGSSWAI--TGVSAA 700	
Db	600 TAATF-TVTTLGSSTTANAI-RAQPSLAGDVWVSTDTGLFHSYNYCKSFQISGCGCTEG 657	
Qy	701 VNVGFGKSPGSSYPAVFVVGITGGVTGAYRSDDCGTTWVLINDDOHOYGN-WGQAITGD 759	
Db	658 WSPFGFGSGDGYPLVFGPFTVDGVTGLFKTEDQGVWQIISDAEHGFGSASANVVNGD 717	
Qy	760 HANLRVYICTNGRGIVYGDIGCAPSGSPSPSVSPSPSPSPSPSPSPSPSPSPSPSPSS 819	
Db	718 LQNYGRVFGTNGRGIFYGD----PSGT-LPSATATASSASTAVKSSSTSTSTSKVGSST 772	
Qy	820 PSSSP 869	
Db	773 TVSSSTATTITSSIKSTLTITTTKSSSTTSSTSTATGTASAYGCGGSGGTGPTQCP- 831	
Qy	870 SGGVKVQYKNN-----DSAPG 885	
Db	832 -SGWTCYENIYSQCKSIPIG 851	

```

RESULT 6
Q9P4T8 PRELIMINARY; PRT; 806 AA.
AC Q9P4T8;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE CEU6 protein precursor.
GN CEU6.
OS Agaricus bisporus (Common mushroom).
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC Agaricales; Agaricaceae; Agaricus.
OX NCBI_Taxid=5341;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D649;
RA Morales-Almora P., Thurston C.F.;
RT "Molecular analysis of the cellulolytic genes in Agaricus bisporus.";
RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ292529; CAC02964.1; -.
DR HSSP; P00725; 2CBH.
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . . ; IEA.
DR GO; GO:0005375; P:carbohydrate metabolism; IEA.
DR InterPro; IPR00254; CBD_fungal.
DR InterPro; IPR002860; GH_BNR.
DR Pfam; PF00734; CBM 1; 1.
DR SMART; SM00236; fCB1; 1.
DR PROSITE; PS00562; CBD_FUNGAL; 1.
KW Signal.
FT SIGNAL
FT CHAIN
FT SEQUENCE
SQ SEQUENCE 806 AA; 84779 MW; 7C67E382E62FF341 CRC64;

Query Match 34.78; Score 1784; DB 3; Length 806;
Best Local Similarity 43.78; Pred. No. 1.2e-76;
Matches 355; Conservative 131; Mismatches 277; Indels 50; Gaps 13;

QY 20 LAATASPAVAALGLVPLAIATASPAHAATTPQYTSNVAI-GGGGFGVDGIVFNEGAPGIL 78
DB 1 MPSEVRFWITLTLTGTSLVY-----GQAASQSYRQNVKIGGGGFGVGFNPSEKGLA 55

QY 79 YRTDIGMYRMDAANGRWIPLLDWGWNNWYGVVSIADPTNTNKKWAAVGMVYNSW 138
DB 56 YARTDIGGAYKLN-ADDTWTPLDLDFAADDRWNTWGVDALATDPVQPNRLATGLYNSW 114

QY 139 DPNDGAILRSSDQATWQITPLPKLGNMPPGMRGERLAVDNDNLIYFGAPSGKGLW 198
DB 115 DPNSGHILISSDYGKTFDAAPLPKLGNMPPGMRGERLVDPNLNLIYFGARSGNGLW 174

QY 199 RSTDGSGATSQMTNFPDVGTYIANLPDITTTGYQSDIQGVVWVAFDKSSSSLGQASKTIFVG 258
DB 175 KSTNSGRSSKSVTSFTDTGTFVDPDSDTGLNSDKIGIAWTTIDKASGSSSATPRIFVG 234

QY 259 VADP-NNPFWSRDGGATWQAVPGATGTFPHKGFPDPVNVHLYIATNTGGPYDSSGD 317
DB 235 VANKGSNIYISTNGSSWSAVAGQPTSTFLPHKGVLSPSERALYITYSDGGGYPDGTSGA 294

QY 318 VKFVSTVGTWTRISVPVPSPTDANDYFGVYSGGLTIDRQHPNTIIVATQISWMPDTIIFRST 377
DB 295 VYKISIDTGAWTDITPVSGSDL---FPFGGLALDQKNGTVMVAALNSWMPDQIIFRST 351

QY 378 DGGATWTRIDWTSYPNRSLEYVLDISAEFWLTFGQPNPPVPSP---KLGWMDAEMAI 434
DB 352 NGCASWTPLDWAVYPTLNKYSYNAALAPW---ICPNIVDVTPTGNLQIGWMMESLSID 407

QY 435 PFNSDRMLYGTGATLYATNDLTWDSGQHIAPMVKGLBEETAVNDLISPPSGAPLISAL 494
DB 408 PFDNSHWLYGTGTIYIGSRDLLKWDASHNVYTIKSLADGVEETSQVALISPPSGPLVSAI 467

QY 495 GDLGGFTHADVAVPSTIFTSVPYFTTGTSDVYAEALNPSIIVRAGSDPPSPQNDRVAFS 554

```

```

Db 468 GDVGGFVHSLTRAPSAQFTNPKNWSTADIDFAGNPNPNVRIIGTGDST--GKQVALS 524
QY 555 TDGCKNWFQSEPGVVTGCTVAASDGRFVWAPGDPQGVVYVYVGFNGSMAASQGVPA 614
Db 525 SDVGVVWQHFGAPDNVQGGKVALSADADILLWRITGNG-----563
QY 615 NAQIRSDRVNPKTFYALSNQTFYRSTDDGVTFOFVAAGLPSGAGVGMFHAVPGKGDLM 674
Db 564 -VMVSRNQATFNIFYGASGKTFYVSTDMGKTFSAHGLSGLSATSVDITVH--PSVSGDIW 620
QY 675 LAASSGLYHSTGSGWSAITGVSSAVNVCFKSAFGSSYPFAVVVCTIGVTCVAVESDD 734
Db 621 ASTDKGLFHSSTGATFSASIGITQAWGVALGAPRSTGGYPVFAAANYGVEVAFESDD 680
QY 735 CGTWTWLINDDOHQYG-NWQOATGDHANLRVYVIGTNGRGIYVGDIGGAPSGSPSPSVS 793
Db 681 RGVNWKINDAAGFGAASANCWAADPRVYGRVYIGTNGRGIYVGDVAGS---APPTSS 737
QY 794 PSAPSLSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 826
Db 738 ATSTTSVSTTTTTRSTTTTTRSTTTTTS-SGNGSP 769

RESULT 7
O74170 PRELIMINARY; PRT; 856 AA.
AC O74170;
DT 01-NOV-1998 (T-EMBLrel. 08, Created)
DT 01-NOV-1998 (T-EMBLrel. 08, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE Avicelase III.
OS AVIII.
GN Aspergillus aculeatus.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurociates; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5053;
RN [1]
RP SEQUENCE FROM N.A.
RA Arai M., Takada G., Kawaguchi T., Sumitani J.;
RT "Avicelase III from Aspergillus aculeatus.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB015511; BAA29031.1; -.
DR PIR; T00349; T00349.
DR HSP; P00725; 2CBH.
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR00254; C:BD fungal.
DR Pfam; PF02012; BNR; 7.
DR Pfam; PF00734; CBM_1; 1.
DR ProDom; PD001821; C:BD fungal; 1.
DR SMART; SM00236; fCBD; 1.
DR PROSITE; PS00562; C:BD FUNGAL; 1.
SQ SEQUENCE 856 AA; 89620 MW; BE085983AF60ED76 CRC64;

Query Match 33.9%; Score 1740; DB 3; Length 856;
Best Local Similarity 43.5%; Pred. No. 1.5e-74;
Matches 366; Conservative 139; Mismatches 237; Indels 40; Gaps 18;

QY 30 AALGVLPALITASPAHAATTQPTWNSVNI-GGGGFVDGIVFNEGAPGILYVRTDIGMY 88
Db 4 SSALLCAALLKGLADAASQATWKNVVTGGGGGFTFVFNPSAKVAVARTDIGAY 63
QY 89 RWDAAANGRIPLLDVWG---WNWNGYGVVSIADPINTKNKVAAGVMTNSWDPNDGAI 145
Db 64 RLN-SDDTWTPLMDVWGNDFHWD--GIDALATDPVTDTRVYVAVGVMTNWDPNVGS 119
QY 146 LRSDQGATWQITPLPKLGNMPCGRGRLAVDPNDNLLYFGASGKGLWSTDSGA 205
Db 120 LRSTDQGDWTETTKLPKFGVGNMPCGRGRLAVDPNKNLSILYFGARSGHGLWSTDYGA 179

```

```

QY 206 TWSQMTNPPDVGTYIANPTDTTGYQSDIQGVVWVAFDKSSSSLSGOASKTIIFYGVADPNP 265
Db 180 TWSNVTFTWTGTFQDSSST--YTSQVGLAWVTFDSTSGSSGATPRIFVGVADAGKS 237
QY 266 VFVSRDGGATWQAVPGAPT-GFIPHKGVDFPNVHVLIYATSNITGPGYDSSGDVMKFSVT 324
Db 238 VFKSEDAGATWAWVSGFPOYGFPHKGLVSPBEKTLIYSYANGAGPYDGTNGTVHKYNT 297
QY 325 SGWTRISPVSTDTANDYFOYSGLTIDRQHPNTIMVATQISWPDITIFRSTDGATWT 384
Db 298 SGVWTDISP--TSLASTYGYGGLSVDLQVPGTLMVAALNCWNPDELIFRSTDSGATWS 354
QY 385 RINDWTSYPNRSLYVLDISAEPLWTFGVQPNP-PVSPKLGMDMAIDFNSDEMXY 443
Db 355 PIWEMNGYPSINYYSYDINAPYQDTSTTDQFV--RVGMVVEALAIIDFDSNHWLY 411
QY 444 GTGATLATNDLTWKDSSGQTHIAPVMKGLBEATVNDLISPPSGAPLISALGDLGGFTHA 503
Db 412 GTGLTVGGHDLTNWDSKHNVTVKSLAVGIBEMAVLGLITPPGFPALLSAGVDDGGFYHS 471
QY 504 DVTAVPSTIFTSVFTTTSVDYAEINPSIIVRAGSFPSSQPNDRHVAFTDGGKWFQ 563
Db 472 DLDAAPNQAYHTPTVTGTTGIDYAGNKPFSNIVRSASD--DYPT--LALSNGFGSTWA 526
QY 564 GSEPGVTTGTTVAASADGRFVWAPGDPQVYVAVGFGNSWAASQGVPAQAIRSDRV 623
Db 527 DYAASTSTGTGAVALSADGDTVLLMSSTSGALVSKSQ--TLTAVSSLPAGVIAADKS 583
QY 624 NPKTFYALSNQTFYRSTDDGVTFOFVAAGLPSGAGVGMFHAVPGKGBDMLWLAASGLVH 683
Db 584 DNTVYGGAGAIYVSKNTATSFTKTVS-LGSSTTVNAI-RAHPSIAGDVWASTDKGLWH 641
QY 684 STNGSGSSMAI-TGVSSAVNVGFGKSAFGSSYPFAVVVCTIGVTCVAVESDDCGTTWLI 742
Db 642 STDYGSTFTQIGSGVTAGWSFGKASSTGVSIVYVIGPTIDGAAGLFXSEDAGTNWQVI 701
QY 743 NDDQHQYGNWQOA--ITGDHANLRVYVIGTNGRGIYVGDIGGAPSG-----SPSPSV 792
Db 702 SDASHGFGS-GSANVWNGDLQTYGRVFRGHERPGHLLRQSRPAGRHGDGDDGDTTTSKT 760
QY 793 SPASPSLSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 852
Db 761 STTVSTTLTKTSSASTTSSSTTVKTTTSSSTTSKASSTTTTKTTTSSSGTIATA 820
QY 853 SS 854
Db 821 SA 822

RESULT 8
Q729M8 PRELIMINARY; PRT; 838 AA.
AC Q729M8;
DT 01-OCT-2003 (T-EMBLrel. 25, Created)
DT 01-OCT-2003 (T-EMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Cel74a.
GN CEL74A.
OS Trichoderma reesei (Hypocrea jecorina).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreomycetidae; Hypocreales; Hypocreaeae; Hypocrea.
OX NCBI_TaxID=51453;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=QMea;
RA Foreman P.K., Brown D., Dankmeyer L., Dean R., Diener S.,
RA Dunn-Coleman N.S., Goedegebuur F., Houfek T.D., England G.J.,
RA Kelley A.S., Meerman H.J., Mitchell T., Mitchinson C., Olivares H.A.,
RA Teunissen P.J.M., Yao J., Ward M.;
RT "Transcriptional Regulation of Biomass-Degrading Enzymes in the
RT Filamentous Fungus Trichoderma reesei.";
RL J. Biol. Chem. 0:0-0(2003).
DR EMBL; AY281371; AAP57752.1; -.

```

SQ SEQUENCE 838 AA; 87132 MW; FBE3D08F2A74FA84 CRC64;

Query Match 31.7%; Score 1625.5; DB 3; Length 838;
 Best Local Similarity 41.7%; Pred. No. 3.9e-69;
 Matches 350; Conservative 131; Mismatches 296; Indels 63; Gaps 23;

QY 25 SFVAALGVLPATITASPAHAATQPTWTSNVAL-GGGGFDGIVNEGAPGLIYRTD 83
 DB 4 SRYLALVLGAV-----IPAHAA-----FSWKVKLGCGGFGVPGIIPHPKTKGVAYARTD 53

QY 84 ICGWYWDANGRWIPLLDHW-----GWNNGYGVSVIAADPTNTNKAAGVMTNSWD 139
 DB 54 IGGYRLN-ADDSWTAVTDGIADNAGHNN--GIDAVLDPQDDQKQYAAAGVMTNSWD 109

QY 140 PNDGAILRSDGATQWITLPLPKLGNMPCRGMRGERLAVDPNNNDNLYFCASGKGLMR 199
 DB 110 PSNGAILRSDRGATWSFTNLPLPKVGNMPCRGMRGERLAVDPANSNLIYFCASGKGLWK 169

QY 200 STDGATWSQMTNFPDVGTYIANPTDTTGYOSDIQGVVWVAFKSSSLGQAOKTIFVGV 259
 DB 170 STDGVTFSKVSFTATGTYIPDSGSDGYNKQGLMWVTFDSTSTTGGATSRIFVGT 229

QY 260 ADP-NNPVFWSDGGATQWAVGAPGTFPHKGVDFDNNHVLATSNTPGPDGSSGDV 318
 DB 230 ADNITASVYVSTNAGTWSAVPGQPKYFPKAKLOPAEKALYLTSDGTPGYDGLTGSV 289

QY 319 WKFSVTSQWTRISPVSTDTANDYFCYGLTIDRQHPNTIMVATQISWMPDTIIFRSTD 378
 DB 290 WRDYIAGGTWKDITPVSGSLD---YFGGGLGDLQKPGTLVVASLNSWMPDQLFRSTD 346

QY 379 GGATWTRIDWTSYPNRSLRYLVIDISAEPWLTFG-VQPNPPVPS-----PKLGWMDAMAI 433
 DB 347 SGITWSPWAWASYPTEYYISITPKAPWIKNFDVTSSESFDGLIKELGWMIESLEI 406

QY 434 DPNSDRMLYGTATLYATNDLTKWSDGGQIHIAPMVKGLBETAVNDLISPPSGAPLISA 493
 DB 407 DPTDSNHLVYGTGWTIFGGHDLTNWDRHNVISQSLADGTEBFVQDLASAPGSELLAA 466

QY 494 LGDLGCFTHA--DVTAVPSTIFTSVPVFTTGTSDVVAELNPSIIVRAGSFDPPSQPNDRH 550
 DB 467 VGDNDGFTFASRNDLGTSPQVWATPTWATSTSDVYAGNSVKSVVRVGTAGTQQ----- 521

QY 551 VAFSTDGGKNNVFOGSEPGGVTTCGTVAASADGSRFVWAPDGPQVYVAVFGNSWAAQ 610
 DB 522 VALSSDGGATSDIYAADTSMNGTWAYSDAGDTILWSTASSG---VQRSQFOGSPASVS 578

QY 611 GVPANAQISDRVNPKTFFVALSNGTYRSTGVTQFQVNAAGLPSGAGVW--FHAVPG 668
 DB 579 SLPAGAVIASDKKTNVFFVAGSGSTFVYKDTGSSF---TRG-PKLGSAGTIRDIAAHT 634

QY 669 KEGDLWLAASSGLYHSTNGSSWSAI-TGVSSAVNVGFGKSARSSYPVAVVGTIGVT 727
 DB 635 TAGTLVSTVDGIFRSTDSGTTFGQVSTALTNTYQIALGVGS-GSNW-NLYAFGT--GPS 690

QY 728 GA--YRSDDCGTTWLINDDQHQYGNWGOAITGDHANLRVYIGTNGRGTVY--GDIGGA 783
 DB 691 GARLYASGDSASWTDIQSQGREGSDSTKVASGSTAGVYVGTNGRGVYAGTIVGGG 750

QY 784 PGSPSPSPSPSASPLSP 843
 DB 751 TGGT-----SSSTKQSSSTSSASSTTLRSSVSTTRASTVTSSTSSAAGPTGS 801

RESULT 9
 Q82K30 ID PRELIMINARY; PRT; 739 AA.
 AC Q82K30
 DT 01-JUN-2003 (TEMBLrel. 24, Created)
 DT 01-JUN-2003 (TEMBLrel. 24, Last sequence update)
 DE 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
 DE Putative glycosyl hydrolase.
 GN SAV2574.
 OS Streptomyces avermitilis.

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycineae; Streptomycetaceae; Streptomyces.
 RN NCBI TaxID=33903;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
 RX MEDLINE=21477403; PubMed=11572948;
 RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
 RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Oonoe T.,
 RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
 RA "Genome sequence of an industrial microorganism Streptomyces
 RT avermitilis: deducing the ability of producing secondary
 RT metabolites";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220 (2001).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN-MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
 RX MEDLINE=22609306; PubMed=12692562;
 RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
 RA Sakaki Y., Hattori M., Omura S.;
 RA "Complete genome sequence and comparative analysis of the industrial
 RT microorganism Streptomyces avermitilis";
 RL Nat. Biotechnol. 21:526-531 (2003).
 DR EMBL; AP005031; BAC70285.1; -;
 DR GO; GO:0016787; F:hydrolase activity; IEA.
 DR InterPro; IPR002860; GH_BNR.
 DR Pfam; PF02012; BNR; 9.
 DR TIGRFAMs; TIGR01409; TAT_signal_seq; 1.
 DR Hydrolase; Complete proteome.
 KW SEQUENCE 739 AA; 78399 MW; 8E3B0F3D0F5B3D28 CRC64;

Query Match 28.1%; Score 1442; DB 16; Length 739;
 Best Local Similarity 40.8%; Pred. No. 1.6e-60;
 Matches 328; Conservative 107; Mismatches 272; Indels 96; Gaps 25;

QY 7 IRLTMRRLSLLAATASFAVAALGVLPATITASPAHAAT---QPTWNSNVAIGGG 63
 DB 1 MRPSPSRRTV--LAGTAA---AAALTAVP-AVCSQAHAETTAGPSYRWNAVIGGTG 53

QY 64 FVDGIVNEGAPGLIYRTDTCGYMYRDAANGRWIPLLDVGNWNNWNGVYVSIADPN 123
 DB 54 FVTGVLFHPSVRGLAYARTDTCGAYRWDRGARWTPLDHLGWDNDWLLGVAMAVDP 113

QY 124 TNKWAAGVMTNSWDNDGAILRSDGATQWITLPLPKLGNMPCRGMRGERLAVDP 183
 DB 114 PDRLYLAVGTVAQSWAGN-GAVLRSEDRGATWTRTDLTVKLGNEEDRGAGERLLVD 172

QY 184 DNILYFGAPSGKGLWRSTDGATWSQMTNFPDVGTYIANPTDTTGYSDIQGVVWVAFDK 243
 DB 173 SDTLWLGT-RHDGLLKSTDGATWAAATAPP-----AKANSSGQGVVF----- 214

QY 244 SSSSLGQAOKTIFGVADPNP-----VFWSRDGGATQWAVGAPTGF---IPHKGVFDP 295
 DB 215 ----LVAAGRVTYAGWGDGCTSGTANLYRTAD-GTTWGAVPGRPSGTSKAPVPLAA 269

QY 296 VNHVLYATSNTPGPDGSSGDVWKFVSTGWTTRISPV-----PSTDANDYFCYGLTI 351
 DB 270 HTRELYTYGDAPGQGGSDGSHVHLRTATGTATGTEVTPKPGGTGSDGADTFAYGG 329

QY 352 DROHPNTIMVATQISWNPDTTIFRSTDCGATWTRIDWNTSYPNRSLRYLVIDISAEP 411
 DB 330 DARRPGTLVSTNNRWADGTVFRSTDCGRTWTSKLD-----AAVFDVSETFPLDW 380

QY 412 GVQPNPVPSPKLGWMDAMAIIDPFNSDRMLYGTATLYATNDLTKWSDGGQIHIAPMVK 471
 DB 381 GDD-----KPKFCWMIQALAVDPYDSQHVYVGTGATLYGTRDLKRW-----APR 426

QY 472 GLEETAVNDLISPPSG-APLISALGDLGGFTHADVAVPST-IFTSVPVTTGTSVDYABL 529
 DB 427 GLEESAVRQLISPPVGEAHLISGLDIGVMYHERLTASPSRGMATNPVFGSATGLAQAAA 486

QY 530 NPSTIIVRAGSFDPPSQPNDRHVAFPSTDGGKWN--FQSGEPGGVTTGGTVAASADGSRFW 587

```

147 RSSDQATQITLPLPKLGGNMPGCMGERLAVDPNNNDNIIYFGAPSGKGLWRSTDGAT 206
141 RSPDRGQFERADLPFKLGGNQLGRANGERLAVDPHGRVILLGRDA-GLWRSDRGAAH 199
207 WSMOTNFPDVGTYIANPTDTTGYQSDIQGVVWVAFDKSSSSISLGQASKTIFVGVADPNP 266
200 WARVEGFPADALAGATARNHVGSQAV-GIAFVVFDAASGHAGATPRIYGVSTAQTS 258
267 FWSRDGATQWAVGAPGTGFIPIHKGVPDPVNHV-----YIATNTGPGPYDSSGDV 319
259 YVSDAGRTWSAVAGQPKGLRP-----SHWGHNAQWYLSYGDPRPGPDLMAGGALW 310
320 KFSVTSGTWTRISVPSTDTFANDYEGYSLGIDRQHPNTIMVATQISHWPDIIIRSTDG 379
311 NYDATQGRWREISPIQPAT-GDGFGWGAVADPQHPQVILLASTPRRTRPRDEVFRSGD 369
380 GATWRIWDTSYPNRSLRYVLDISABPWLFGVQPNPVPSPKLGWMDMAAIDPFNSD 439
370 GRSWVPLAAAPF-----DHSAAPW-TAHATPH-----WIG-ALAIIDPFDSN 409
440 RMLYGTGATLYATNDLTKMDSGGQIHIAPMVKLEETA VNDLISPPSGAPLISALDGLG 499
410 HATFVTGVIWASRNLTFSQQPLQWFWQDRGLEETVPLDLSPMAGAHLLSALGDIDG 469
500 PTHADVTAVPSTIFTSPVTTGTSVDYAEALNPSIIVRAGSPDPSSQPNDRHVAFSTDGK 559
470 FRH-DALDTAQLQYLGPRLTNGESIDGAGQAPQWVYRSQTV-RDRNNEIRALYSQDGT 527
560 NWFQ-GSEPPGGVTGTVAAADSGSRFVWAPGDPQPVVYAVG-FGNSWAAASQGVPAQA 617
528 HWAFAFSEPPRGGAGTITAIADASQVWVPDQGG---VMRTGDFGKRWQVQGLPDTAV 584
618 IRSRVNPKTFYALS--NGTFYRSTDDGVTFOFVA--AGLPSSGAVGVNMFHVPKKEGDL 673
585 VVADRVDAQRWYAADRVSGRLYESDGAASPRDTGQGVSGPARDE-----RARPQLRDP 639
674 WLAASSGLYHSTNGSSWS-----AITGVSSANVGVFGKSPAGSSYPVAVVVTGTTGVTG 728
640 WRAGVYVLAASPTLGMVWQNGQLRTLKSDARSIGIKALRAGAPPALYLAGRVAGVDG 699
729 AYRSDDCGTTWVLINDDOHOGNMGQAITGHDANLRRVYVIGTNGRGIVYGD 779
700 IFRSDDEGAHRRINDDAHFRGK-PYSVTGDPRAGRVTFATGGRGIFYGD 749

RESULT 11
Q9WYE1 PRELIMINARY; PRT; 707 AA.
AC Q9WYE1;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Endoglucanase, putative.
GN TM0305.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=9287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickley E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
genome sequence of Thermotoga maritima."
RL Nature 393:323-329(1999).
DR EMBL; AB001712; AAD35393.1; -.
DR PIR; F72393; F72393.
DR TIGR; TM0305; -.

```

```

487 RPAVVRTGWGDHNG-----AYSHDGGRTWAPFAQPDIAKADGPIATSDGGTLLW 540
588 A-----PGDPCQPVVYA-----VGFENSWAASQGVFANAAQIRSDRVNPKTFYA--LSNGTFY 637
541 SFVHWGD-----TTAAHRSTNDGASWSESVSPFKGATPVADPPTFRFVAYDFDNGTLY 595
638 RSTDGVTQFVVAAGLPPSSGAVGMFHAVPKGGLWLAAAS-SGLYHSTNGGSSWSAIGT 696
596 ASTDSGRSEFARAGLP-SGDSQFKLVAAPGRSGDLWLSAKMNGLYRSTDGGDTFARIDS 654
697 VSSAVNVGSKSAPGSSYPVAVFVGTGIGVAYRSDDCGTTWVLINDDOHOGNMGQAI 756
655 CWASYTLGFGKAADGADYPAIYQVSTETTAYVRSDDAARTWVRINDDAHQWNGICEAV 714
757 TGDHANLRRVYIGTNGRGIVYGD 779
715 VGDPRHGRVYLATNGRGIOYGE 737

RESULT 10
Q8PLM5 PRELIMINARY; PRT; 751 AA.
AC Q8PLM5;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Cellulase.
GN CELA OR XAC1770.
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=92829;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=306 / ATCC 13902 / XV 101;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Cannavan F., Cardoso J., Chamberg J., Ciapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locati E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spindola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Tindade dos Santos M., Truffi D., Tsai S.M., White P.P.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
host specificities."
RL Nature 417:459-463(2002).
DR EMBL; AE011809; AAM36634.1; -.
DR InterPro; IPR002860; GH_BNR.
DR Pfam; PF02012; BNR; 7.
KW Complete proteome.
SQ SEQUENCE 751 AA; 81259 MW; D4134D96285E599D CRC64;
Query Match 22.9%; Score 1173.5; DB 16; Length 751;
Best Local Similarity 35.5%; Pred. No. 8.1e-48;
Matches 274; Conservative 117; Mismatches 319; Indels 61; Gaps 20;

```

```

27 AVAAALGVLPITATSPAHAAATQPTWYSNVAIGGGGFGVDFVNEGAGILYVTRDIG 86
22 AVCLTLLVFWTSEASDEPTDPPYQWHSVAIGGGGFGVDFVLPFAERDLAYARTDVG 81
87 MYRWDAANGRWIPLLDVWGNNGYGVSAADPINTNKWAAVGYTNSWDNDGAIL 146
82 AYRWDAQOQWALTDWLGADDDNLMGIDAFVADPADPNALYLAAGTYMHERAGN-AVL 140

```



```

Qy 491 ISALGDLGGFTADVTAVPSTTFTSPTVTTGTSVDYAEALNPSIIVRAGSPDPSSQPNDRH 550
Db 461 LSGAGLIDGIDRDELDR-OLQVAGPRLTNGESIDAAGQAPQWVRSGTV-RDRRNEIR 518
Qy 551 VAFSTDDGKNWQ-GSEPGVTTGGTVAASADSGRFVWAPGDPQPVVYAVGFGNSWAAS 609
Db 519 ALYSRGGQKWTFAPASEPPAGQAGSSTAIGADAQVVMAPERGN--WRTSDFGAQWQV 576
Qy 610 QGVPAQAQIRSRVNPKEFYA--LSNGTFYRSTGGVTTFQVAAGL----- 653
Db 577 DGLPDTAVVMADEVDAARWYADVASQLYESTDAARSFR--ATGVQVGSFARDERTRPQ 634
Qy 654 --PSSGAVGMFHAVPGKEGDLMLAASSGLYHSTNGSSSAITGVSSAVNNGFGKAGP 711
Db 635 LRDPWRAGVVYLASPGK-----GVMRQDG--TLQVLSQPDPEARSLGTGKALRA 682
Qy 712 SSVPAVAVVGTGGTGVAYESDDGTTWVLINDDOHQYGNWGAQITGDHANLRVYIGTN 771
Db 683 GAPPALYLAGRVQGVGVFRSDDGQVQWQORINDDAHRFR-PYSVTGDPRIAGRVYATG 741
Qy 772 GRGIVYGD 779
Db 742 GRGIFYGD 749

RESULT 13
Q8J0D2 PRELIMINARY; PRT; 812 AA.
ID Q8J0D2 PRELIMINARY; PRT; 812 AA.
AC Q8J0D2;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Oligoxylucan reducing end-specific cellobiohydrolase.
OS Geotrichum sp. M128.
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Dipodascaceae; mitosporic Dipodascaceae;
OC Geotrichum.
OX NCBI_TaxID=203496;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M128;
RX MEDLINE=22359052; PubMed=12374797;
RA Yaci K., Mitsuishi Y.;
RT "Purification, Characterization, Cloning, and Expression of a Novel
RT Xylolucan-specific Glycosidase, Oligoxylucan Reducing End-specific
RT Cellobiohydrolase.";
RL J. Biol. Chem. 277:48276-48281(2002).
DR EMBL; AB089343; BAC22065.1; -.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR InterPro; IPR002860; GH_BNR.
DR Pfam; PF02012; BNR; 9.
KW Hydrolase.
SQ SEQUENCE. 812 AA; 87080 MW; F5C740F32D8B17B5 CRC64;

Query Match 21.9%; Score 1125; DB 3; Length 812;
Best Local Similarity 34.5%; Pred. No. 1.7e-45;
Matches 285; Conservative 129; Mismatches 291; Indels 120; Gaps 31;

Qy 31 ALGVLPATASPAHAATTPQYWSNVAIGGGVFDGIVNEGAPGILYVRTIDIGMYRW 90
Db 6 SLKALTALSILASLAVAKHEHYEFKNAIGGGVITGIVAHPTKOLLVARTIDIGAYRW 65
Qy 91 DAANGRWIPLLDVGNWNNYGVVSTAADPINTKVAAGVMT-NSWDPNDGAILRSS 149
Db 66 DACTSKWIPUNDIEADNMINGTESIALDPNPNDRLLYLAQGRVYGDW-----AAFYVSE 121
Qy 150 DQATWQITPLPKLAGNMPGRGGERLAVDPNNDNLYFCAPSGKGLWRSTDSGATWSQ 209
Db 122 DRQGSTIYESPPMGANDGRNGERLANVNPNSNEVWGTGT-EGIKSSDRAKTWTN 180
Qy 210 MTNFPVGVTVIANPTDTTGYQSDIQGVVWVAFDKSSSLGQASKTIFVGVADPNPNFWS 269
Db 181 VTSIPDAF-----TNGIGYTS-----VIFDP-----ERNGTIVASATAPQG-MYVT 220

```

```

Qy 270 RDGATWQAVPGAPTGI-----PHKGVDPVNVHVLVIATSNVTGGPY 311
Db 221 HDGVSWEVPVAGQPSWLNRTTGAPDPKPAASIAQPQMKVALTP--NFLYTVYADYFGEW 278
Qy 312 DGSSEGVWKSFTVSGTWTRI-----SPVSTDTANDYFGYSGLTIDRQHNPTIMVATQ 364
Db 279 GVTGFEVWRQRTSGAWDDITPRVGNSSPAPNNQTFPAGFCGLSDVATNPNRLVVIIT- 337
Qy 365 ISWMPDRI---IFRSTDCGATWTRIWDWTS-----YFNRSLRYVLIDISAEPLWTFG 412
Db 338 LDRDPGALDSIYLSSTDAGATWQDVQLSSPSNLEGNWGHPTNAARY-KDGPVVPWLDFN 396
Qy 413 VQP-----NPVPSP---KLGWDEMAAIDPNSDRMLYGTGATLYATNDLTK-----WDS 460
Db 397 NGPQWGGYGAPHGTPGLTKFGWMSAVLIDPPNPEHLMYGTGATIWATDLSRVEKDW-- 454
Qy 461 GGQIHIAP---MWKGLEETAVNDLISPPSGAPLISALGDLGFGTHADVATVPSTIFTSP 516
Db 455 -----APSWYLIQDIGEENAILSLRSPKSGAALLSGIGDISGMRHDDLTK-PQKMFAP 507
Qy 517 VFTTGSTVDYAEALNPSIIVRAGSPDPSSQPNDRHVAFTSDGGKNW--FGSPPGGVTT-- 572
Db 508 QFSNLDSDAAGNPNVVRAGSSGHEYDSACARGAYATDGGDAWTFITCTCPGMWASHY 567
Qy 573 -GGTVAASADGSRFVWAP--GDPGQPVVYAVGFGNSWAASQGVPA-----NAQIRSDRV 623
Db 568 QGSTIAVDASGSIQVWSTKLDQASGPMWYSHDYGKTSW---VPAGDLKAQATANVLSDKV 623
Qy 624 NPKTFYALNSGTFVRSSTGGVTFQVVAAGLPSGAGVGMFHAVPGKEGDLWL-AASSGLY 682
Db 624 QDGTFYATDGGKFFVSTDDGGSYAAKAGLVLT--QTSLMPAVNPWVAGDVMVPEGGGLF 681
Qy 683 HSTWGGSGSWSAI-TGVSAVNVGFKS---APGSSYPVAFVVGVT--IGCVTGAYRSDDC 735
Db 682 HSTDFGASFTRVGTANATLVSVGAPKSKSDGKASAPSAVFVGTGDKPSDGLYRSDDN 741
Qy 736 GTTWVLINDDOHQYGNWGAQITGDHANLRVYIGTNGRIGVYGD 780
Db 742 GSTWTRVNDQEHNYSG-PTMIEADPKVYGRVYLGTNGRIGVYADL 785

RESULT 14
O52374 PRELIMINARY; PRT; 1779 AA.
ID O52374;
AC O52374;
DT 01-JUN-1998 (TRENBLrel. 06, Created)
DT 01-JUN-1998 (TRENBLrel. 06, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Family 10 xylanase (EC 3.2.1.8).
GN XYN.
OS Caldicellulosiruptor sp. Rt69B.1.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=70295;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rt69B.1;
RA Morris D.B., Gibbs M.D., Ford M., Thomas J., Bergquist P.L.;
RT "Family 10 and 11 xylanase genes from Caldicellulosiruptor sp.
RT Rt69B.1.";
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY F (FAMILY 10 OF GLYCOSYL
CC HYDROLASES).
DR EMBL; AF036924; AAB95326.1; -.
DR PIR; T31085; T31085.
DR HSSP; Q06851; INBC.
DR GO; GO:0030246; F:carbohydrate binding; IEA.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; F:carbohydrate metabolism; IEA.
DR InterPro; IPR001956; CBD 3.
DR InterPro; IPR00584; CBD IV.
DR InterPro; IPR005084; CBM_6.

```

DR InterPro; IPR003305; CEM_Cenc.
DR InterPro; IPR008965; Cellul. bind.
DR InterPro; IPR008979; Gal. bind. like.
DR InterPro; IPR001000; Glyco_hydro_10.
DR Pfam; PF00942; CEM_3; 3.
DR Pfam; PF02018; CEM_4; 9; 2.
DR Pfam; PF03422; CEM_6; 1.
DR Pfam; PF00331; Glyco_hydro_10; 1.
DR PRINTS; PD00134; GLHYDRLASE10.
DR ProDom; PD001947; CBD_3; 3.
DR SMART; SM00606; CBD_IV; 1.
DR SMART; SM00633; Glyco_10; 1.
DR PROSITE; PS00591; GLYCOSYL_HYDROL_F10; 1.
KW Glycosidase; Hydrolase; Xylan degradation.
SQ SEQUENCE 1779 AA; 194304 MW; CE5269B6806B5CED CRC64;

Query Match 8.4%; Score 431.5; DB 2; Length 1779;
Best Local Similarity 19.8%; Pred. No. 2.7e-12;
Matches 252; Conservative 162; Mismatches 382; Indels 477; Gaps 53;

QY 65 VDGIVFNEGAPGILYV-----RTDIGMYRWDANGRIPLLDWGMNNGYNGVVSIAAD 120
DB 20 VGTILFHOEAKAAAYVDPECTDTLSFAFKSNIA-----VDMGN-AYNGKSSIRVS 71
QY 121 PINTNKVMAAVGMVTSNDPNDGAILRSSDQATWQIT-----PLPFKL-----G 165
DB 72 --NRSSIDGAVV-----DVKNIMNGTTWVVSAYVKHSYQKPVAFGISAVYDDG 119
QY 166 GNPGRGMRGLAVDPNDNILYFGAPSGKGLWRSTDSGATWSQNTWPPVGTIAPPTD 225
DB 120 SGVKSTLIGEVAI-FN-----YWKXI-----VGRKTPNISN 150
QY 226 TTGQSDIQGVVWVAFDKSSSLQASKTIFVGADPN-----NPVFWSR---DGGAT--WQ 277
DB 151 VRNLLIVVHTVSGVDYND-----YIQMDNYSLSNAVTFSSGFESGTEGQW 201
QY 278 A----VPGAPTGFPHKGVDPNVHVLIA--TSNTGG---PYDG---SSGDVWKFVS--- 323
DB 202 ARGSGVTVPDSVYANGKYS-----LYSGRTSNHGAQIPVDITILEQGVKYSVWVY 256
QY 324 -TSGTWTTRISFVSTDPAND-YFGYSGLTIDRQHPNTIMVATQISWMPDTIIFRSTGGA 381
DB 257 QNSGSTQKMSLTQRRFATDPSTSYENLYNRDVPSTWNTVEPSGSY-----SIPAGV 308
QY 382 TWTIRMDWTSYPNRSRY-----VLDIS--AEP----- 407
DB 309 TVSELLEYBEAQANLAFWVDLKIYDLSKLAPEWEIPSLIEKYDYKFKGVALSYSKI 368
QY 408 -----WLTFGVQPNPPVPSPKLGW 426
DB 369 ASDTEKMWLKHFNISITAGNEMKPSSELLISENNYNFKADEFVNFAFNSNIAIRGHTLV 428
QY 427 MDEA-----MAIDP----- 435
DB 429 HEQTPDWFKANGNTLSKALLSRLLKQYITVTVVGRYKGYAVMDVNEAIDESQGNFR 488
QY 436 -----FNSDRMLYGTGATLYA-----TNDLTQWDSGGQIHAPWKGLEFANVLD 481
DB 489 RSNWYNICGPEYIEKAFIWAHEADPDAKLFYNDYNTENSQKRFYNNIKSLKE----- 542
QY 482 ISPPSGAPLISALGDLGGFTHADVTPSTFTSPVFTT--GTSVDVYAEINPSIIVRAGS 539
DB 543 ----KGVP-IHGIG-LOCHINLWPSISENTIKLFSIPGLEIHITELDMSFYQWGS 596
QY 540 FDPSSQPN-----RHVAF-----STDGKN-----WFO 563
DB 597 TSYSTPRDLLIKQAMRYKELFDLKKYNYITNTVFWGLKDDYSWLSQNFCKSDYPLLED 656
QY 564 GS-----EPGVTGGTV----- 580
DB 657 GNYKSKYAFWSLIEPTVVPVNSTLPAPPAIQVPTPTPTPTPTPTPTPTPTPTPTPTPT 716
QY 581 DGSRFVWAPGDPQPVVYAVFGNNSMAASQGVAPNAQIRS-----DRVNP 625

DB 717 GGS--YWTPE-----SYGALKVWYANGNWSSTTNVLNPKIKIENVGTTAVDLGRKV 767
QY 626 KTFVALSNGTFYRSTGDTGVTQP-----VAAGLPSSCA-----VGMFHAVP 667
DB 768 RYWTYI-DGEAAQSVSVASSINPAIDVVRVVKLGANAGGADYYVEVFGKSGAGVLAAGQS 826
QY 668 QKEGDLWLAASGLYHSTNGSSWSA-----ITGVSSAVNVGFGKAPGSSYPAVFV 719
DB 827 TKEIRLSIQKSSGYNQNDYSVRGSANSYIENEKVTGYIDDLVLMGREGPGRNAQIKWYA 886
QY 720 VGTIGGVTA-----YRSDDCGTT-----WVLINDDOHOYGNWQQAITGDHANLR 764
DB 887 NGNLSSPTNLNPKIKIENVGTTAVDLGRKVRYWYITIDGEATQSVSVASSINPAYIDVR 946
QY 765 RVIYGTNGRGIVY-----GDIGCA-----PSGS----- 787
DB 947 VVKLGANAGGADYYVEVFGKSGAGVLAAGQSTKEIRLSIQKSSGYNQNDYSVRGSANSY 1006
QY 788 -----PPFSVSPASPSLSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 828
DB 1007 IENEKVTGYIDDLVLMGKEPSRGTKPAG--EVTAPTPTSTPTPTPTPTPTPTPTPTPTPT 1064
QY 829 SP 883
DB 1065 TVTATPT 1124
QY 884 PCDNIQKPLQVNVNTGSSVDLSTVTVVYVWTRDGGSSLTVYNCWAAICGCGNIRASFGS 943
DB 1125 ASTGSIRFWKLVNGSSVDLSRVKIRYWTVDGKPOSV-CDWAAQIGASNVTFNFVK 1183
QY 944 VNPATPTADTYLQ 956
DB 1184 LTSGVSGADYILE 1196

RESULT 15
Q8JLH7 PRELIMINARY; PRT; 196 AA.
ID Q8JLH7
AC Q8JLH7
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Putative endoglucanase (Fragment).
GN EGL.
OS Agaricus bisporus (Common mushroom).
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC Agaricales; Agaricaceae; Agaricus.
OX NCBI_TaxID=5341;
RN [1]
RP SEQUENCE FROM N.A.
RA Morales-Almora P.;
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RA Morales P.;
RT "Molecular analysis of the cellulolytic genes in Agaricus bisporus.";
RL Thesis (2001), Department of Microbiology, University of London,
RL London, United Kingdom.
DR EMBL; AJ534359; CAD58875.1; --
FT NON_TER 1 196
FT NON_TER 1 196
SQ SEQUENCE 196 AA; 21353 MW; D9308B2B0935FA7 CRC64;

Query Match 8.0%; Score 409; DB 3; Length 196;
Best Local Similarity 45.5%; Pred. No. 2.5e-12;
Matches 76; Conservative 30; Mismatches 51; Indels 10; Gaps 3;

QY 320 KFSVTSGTWTTRISFVSTDTTANDYFGYSGLTIDRQHPNTIMVATQISWMPDTIIFRSTG 379
DB 1 KYSIDTGAWTIDITPVSGSDL---PFGFGALDITQKNGTVMVAALNSMWDGQIFRSTNG 57
QY 380 GATWTRIDWTSYPNRSRYVLDISAEPLWITFGVQPNPPVPSP---KLGWMEAMAIIDPF 436

Db	58	GASWTLPLWDWAVTTLNKYYSYNAAPAPW----	IGPNIVDVTPGNLQIGWMESLSIDPF	113
Qy	437	NSDRMLYGTGATLYATNDLTKWDSGGQIHAPMVKGLEETAVNDLIS	483	
Db	114	DSNHLYGTGETIYGSRDLLKWDSAHNVTIKSLADGVEETSQUALIS	160	

Search completed: May 14, 2004, 09:18:01
Job time : 55 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 14, 2004, 09:15:08 ; Search time 23 Seconds
(without alignments)
2148.089 Million cell updates/sec

Title: US-09-917-376-1

Perfect score: 5135

Sequence: 1 MDRSENIRLWRSRLVSL.....RASFGVNPATPTADTYLOX 957

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/ptodata/2/iaa/5A COMB.pcp.*
- 2: /cgn2_6/ptodata/2/iaa/5B COMB.pcp.*
- 3: /cgn2_6/ptodata/2/iaa/6A COMB.pcp.*
- 4: /cgn2_6/ptodata/2/iaa/6B COMB.pcp.*
- 5: /cgn2_6/ptodata/2/iaa/PCTUS COMB.pcp.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
1	334	6.5	1751	3	US-09-136-574A-44	Sequence 44, Appl
2	301.5	5.9	1749	4	US-09-640-419C-28	Sequence 28, Appl
3	288	5.6	96	4	US-09-119-507B-112	Sequence 112, App
4	288	5.6	96	4	US-09-547-693-236	Sequence 236, App
5	276	5.4	1426	3	US-09-136-574A-43	Sequence 43, Appl
6	260	5.1	2736	4	US-09-252-991A-30227	Sequence 30227, A
7	234	4.6	2137	4	US-09-134-001C-4463	Sequence 4463, Ap
8	225.5	4.4	551	2	US-09-033-537A-1	Sequence 1, Appli
9	216.5	4.2	8991	4	US-08-714-741-32	Sequence 32, Appl
10	215.5	4.2	76	4	US-09-547-693-233	Sequence 233, App
11	213	4.1	3892	4	US-09-328-352-5503	Sequence 5503, Ap
12	210.5	4.1	521	1	US-08-276-213-3	Sequence 3, Appli
13	208.5	4.1	206	4	US-08-529-055-54	Sequence 54, Appl
14	208.5	4.1	493	3	US-09-198-956-10	Sequence 10, Appl
15	208.5	4.1	493	3	US-09-198-955A-12	Sequence 12, Appl
16	208.5	4.1	493	4	US-09-694-531-12	Sequence 12, Appl
17	208.5	4.1	493	4	US-09-670-141-10	Sequence 10, Appl
18	208.5	4.1	493	4	US-10-072-152-12	Sequence 12, Appl
19	205	4.0	918	4	US-09-200-650B-1	Sequence 1, Appli
20	202.5	3.9	476	4	US-09-339-159B-4	Sequence 4, Appli
21	200.5	3.9	700	2	US-07-862-588B-2	Sequence 2, Appli
22	200	3.9	183	4	US-08-529-055-50	Sequence 50, Appl
23	199.5	3.9	490	3	US-09-109-841-2	Sequence 2, Appli
24	199.5	3.9	616	3	US-09-136-574A-47	Sequence 47, Appl
25	197	3.8	1060	4	US-08-911-393-2	Sequence 2, Appli
26	191.5	3.7	412	1	US-08-313-288B-18	Sequence 18, Appl
27	191	3.7	105	4	US-09-547-693-230	Sequence 230, App

28	189	3.7	933	3	US-08-293-728-2	Sequence 2, Appli
29	189	3.7	933	3	US-09-421-868-2	Sequence 2, Appli
30	188.5	3.7	193	4	US-08-529-055-49	Sequence 49, Appl
31	187	3.6	936	4	US-08-956-171E-5249	Sequence 5249, Ap
32	186.5	3.6	104	4	US-09-547-693-235	Sequence 235, App
33	186	3.6	423	2	US-08-760-797A-1	Sequence 1, Appli
34	185.5	3.6	424	3	US-08-932-929B-1	Sequence 1, Appli
35	185.5	3.6	1719	2	US-08-459-568-4	Sequence 4, Appli
36	185.5	3.6	1719	2	US-08-399-411-4	Sequence 4, Appli
37	185.5	3.6	1719	3	US-08-516-859A-4	Sequence 4, Appli
38	185.5	3.6	1719	4	US-09-586-472-4	Sequence 4, Appli
39	185.5	3.6	1719	4	US-09-528-706-4	Sequence 4, Appli
40	185	3.6	334	6	5202236-3	Patent No. 5202236
41	184	3.6	331	6	5202236-37	Patent No. 5202236
42	181	3.5	167	5	PCT-US95-13813-9	Sequence 9, Appli
43	180.5	3.5	424	2	US-08-760-797A-3	Sequence 3, Appli
44	180.5	3.5	424	3	US-08-932-929B-3	Sequence 3, Appli
45	179.5	3.5	1481	2	US-08-616-844-40	Sequence 40, Appl

ALIGNMENTS

RESULT 1

US-09-136-574A-44

; Sequence 44, Application US/09136574A

; Patent No. 6294366

; GENERAL INFORMATION:

; APPLICANT: Farrington, Graham K.

; Anderson, Paige

; Gibbs, Moreland

; Bergquist, Peter

; Daniels, Roy

; Morgan, Hugh W.

; Williams, Diane P.

; TITLE OF INVENTION: Compositions and Methods for

; Treating Cellulose Containing Fabrics Using Truncated

; Cellulase Enzyme Compositions

; NUMBER OF SEQUENCES: 49

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Howson and Howson

; STREET: Spring House Corporate Center, P.O. Box 457

; CITY: Spring House

; STATE: PA

; COUNTRY: USA

; ZIP: 19477

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/136,574A

; FILING DATE: 19-Aug-1998

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/932,571

; FILING DATE: September 19, 1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Bak, Mary E.

; REGISTRATION NUMBER: 31,215

; REFERENCE/DOCKET NUMBER: 1997US001/CIP

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 215-540-9200

; TELEFAX: 215-540-5818

; TELEX: <Unknown>

; INFORMATION FOR SEQ ID NO: 44:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1751 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; SEQUENCE DESCRIPTION: SEQ ID NO: 44:

US-09-136-574A-44

Query Match 6.5%; Score 334; DB 3; Length 1751;
Best Local Similarity 21.2%; Pred. No. 5.6e-13;
Matches 219; Conservative 125; Mismatches 314; Indels 376; Gaps 56;

Qy	18	SLLAATASFAVAALGVLPIAITASPAHAA-----TTQP-----YWSNVAI	59
Db	186	SVVAET-----AASLAVASVVIKERNSSOKAASYLQHAOKLFEFADTTFRSDAGYTAATGFY	240
Qy	60	GGGFFVDGIVFNEGAPGILVVRTD-----IGGMVRW-----DAANGRWIP	99
Db	241	TSGGPFIDDLGW--AAWLLIATNDSSYLTKAEBLMSEYANGTWTQCHDDVRYGTLIM	297
Qy	100	LLDWVGNNWNGYVWSIAADPINTNKVWAAVGM-YTNSMDPNDGAILRSSDQG-----	152
Db	298	LAKITGKEL--YKGAVERNLDHWTDRITYTPKGMAYLTGM---GSLRYATTAFLACVY	351
Qy	153	ATWQ-----ITPLPKLGNMGR---GMGERLAVDNNNNILYFGAPSG	194
Db	352	ADMSGCDNSKTKYLNFAKSOIDYALGST--GRSFVVGFGYNYPQHHRN-----	400
Qy	195	KGLWRSTDSGATWSQMTNFPD-----VGTYIANPTDTTQSDIQGVWV--VAFDKSSS	246
Db	401	-----AHSSWANSMKIPEYHERHILYGALVCGPGSDDSYNDITDYVQNEVACDYNAG	452
Qy	247	SLGQAOKTIFVGVADPNPNVFWSDGDAQWAPGATGTFIPHKGVFPDPVNNHLYIAT--	304
Db	453	IVGALAK-----MYQLYCGEPID--DFKAIETPTNDEIFVESKF	489
Qy	305	SNITGGPYDSSGGDWKFSVTSGTWTRISVPVSTDTAN-DYP-----GYSGLITDRQ	354
Db	490	GNSQGP---NYTEVISIYNRWTGW---PPRVTDKLSFKFIDITELIQAQYS-----	535
Qy	355	HPNTIMVATQISWMPDTIIFRSTDGGATWTRINDWTSYPNRSLRYL-DISAEPLWTFGV	413
Db	536	PDVVKVDT-----YYIEGGKISGPPVMD---KRNIIYVLVDFSGTK-----I	575
Qy	414	QPNPPVPSPKLGW---MDRAMALDPNSDRMLYGTGATLYATNDLTKWDSGGQIHAPM	469
Db	576	YPGGEVHEHKQAQFKISVPGPYMDPTN-DPSYKGLTSQLEKNKYIAAYDNNN-----	628
Qy	470	VKGLEETAVNDLISP-----PSGAPLISALGDLGTFHADTAVPSTIFT---SPVFT	519
Db	629	VWGLEPGCAATSTPAPTSTPTPTPTPTPTVTA-----TPPTPTPTTGPSP--G	674
Qy	520	TGTSVDYAEALNPSIIIVRAGSFPDSSQPNDRHVAFSTDGGKNWFOGSPBGVTTGCTVAAS	579
Db	675	TGSGVKVLYKNNETSASTGIRP-----WPK-----IVNGG--SSS	708
Qy	580	ADGSRF---VW--APGDPGPVYVAVGFGNSWAASQGVAPANAQIRSDRNPKTFYALSNG	634
Db	709	VDLSRVKIRWYTVDDGKPOSACV-----DW-----AQIGASNVT--FNFVKLSSG	752
Qy	635	TFYRSTDGGVTFQPVAAAGLPSGAGVGMFHAPGKE--GDMLWAASSGLYHSTNGSGSSWA	693
Db	753	V-----SGADYY--LEVGF--SSGAGQLQ----PGKDTGDIQVRFNKNDWSNYNQADWSW	800
Qy	694	ITGVSSAANVFGPKSAPGSSYPVAVFVGTGTGAYRSDDCGTTWVLINDDQHOYGNWG	753
Db	801	LQSMTN-----YGENAKVTLY-----VDG-----VLV-----WG	824
Qy	754	QAITGDHANLRVYIGTNGRGIVVGDIGGAPSGSPSVSPSAPSLSPPSPSSSPSPS	813
Db	825	Q-----EPGGA-----TPAPTSTAPT	841
Qy	814	PPSSSSPSSSP	873
Db	842	PTTAFTPTPTPTPTPTVTSATPTPATSPATSPVGSYWTSPSES-----YCAL	887
Qy	874	KVOYKNDSPAGDNQIKPGLQVNTGSSVDLSLTVTVRYMPTROGGSSLTLYNCNDMAAIG	933
Db	888	KWYANGNLSSPTNVLNPKIKIENVGTTAVDLRSVKRYWYTTIDEGATQSV-----	938

QY	934	CGNIRASFGSVNPA	947
		:	
Db	939	----SVASSINPA	947

RESULT 2

```

US-09-640-419C-28
; Sequence 28, Application US/09640419C
; Patent No. 6630615
; GENERAL INFORMATION:
; APPLICANT: Bidney, Dennis L
; APPLICANT: Crasta, Oswald R
; APPLICANT: Hu, Xu
; APPLICANT: Lu, Guihua
; TITLE OF INVENTION: DEFENSE-RELATED SIGNALING GENES AND METHODS OF USE
; FILE REFERENCE: 35718/199009 (5718-92)
; CURRENT APPLICATION NUMBER: US/09/640,419C
; PRIORITY FILING DATE: 2000-08-17
; PRIOR APPLICATION NUMBER: 60/149,656
; PRIORITY FILING DATE: 1999-08-18
; PRIOR APPLICATION NUMBER: 60/206,405
; PRIORITY FILING DATE: 2000-05-23
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 28
; LENGTH: 1749
; TYPE: PRT
; ORGANISM: Synchocystis PCC6803
US-09-640-419C-28

```

Query Match 5.9%; Score 301.5; DB 4; Length 1749;
Best Local Similarity 19.0%; Pred. No. 6.8e-11;
Matches 225; Conservative 117; Mismatches 402; Indels 439; Gaps 46;

Qy	17	VSILAATASTASFAVAALGVLPTAITASPAHAATTPQYTWSNVAIGG-----62
Db	23	LALLACSSFSFG-----NVLAQNITPAPDTGTVDQAQGNQFNFGVSGDQGNLFHSIQ 78
Qy	63	-----GFVDGIVFNEGAPGILYVTRDITGGMWRDA 92
Db	79	QFGLDQGGIANFLSPDTRNLTIRIVGGDASTINGLIQVSGGNANLFLMNPAGMIFGPNA 138
Qy	93	A-----NGRWIPLLDWGNNWNGYGVVS-----116
Db	139	SINVPGEVVTTSAGNAGFNDQFQV-----PSDNDYNALIGNPSQFAFDLANPGLIINA 193
Qy	117	-----TAADPINTKNVAAVGMVYTNWDPNDGAILRSS-----149
Db	194	GDLSTVEGKNLTFELAGNIVNTGSLAAPCGNITVAAPGQNRIRISQAGSLLSLEVEVSPQ 253
Qy	150	-DQATQWQITPLPPLKGGNMPGRGMGERLAVDPNND-----NILYFGAPSGKGLWRSTD 202
Db	254	MNQGGSFVLDPILLTTCQASNLDLG--LAQPNGSVTTNGTNALVSPILPGSVTISGNVD 311
Qy	203	SGATWSQMTNFPDVGTYITANPTDITGYQS--DIQ-----GVVWAFD--KSSSSLCQASK 253
Db	312	AS-----GKSTNISS--GGVVAIAGDQIAVQCATVDVSGNGGGGTVRIGDGFQGOITLFPNASQ 367
Qy	254	TIFVGVAADPNNPV-----FWSRDGGA--TWQAVPGAPTGFIPHKGVFDPVNVHLYIATSNT 307
Db	368	TLI-----DSNSVVRADALLTNGGTVIWWADSTRFSGNI-----SAQ 406
Qy	308	GGPYDGSBGDWKFSVTSGETWTRISPVFSTDANDFYGYSGLITDRQHPNTIMVATQISW 367
Db	407	GGTWGGNG-----FVETSG-----AKSLMVDDETARVNTFATMGELGT 444
Qy	368	W---PDTIIFRSTDCGATWTRWD-----WTSYPNRSRLRYVLDISAEPMLTFPGVQNPVPV 420
Db	445	WLLDPLEIIVGTTDDLLADPKLVSVLVTITTSLDNGNVILQAQSIJAVQANFSADPSAP-- 502
Qy	421	SPKLGWMDENAAIIPFNSDRMLYGTGATLYA-----TNDLTKWDSGGQIHI-- 466

	D	b	503	-GMLTFDSPTIITIDALFS-----LGTSGIIIPANTGPINTGNLTIVTSPFTNLDLDFDKIQIALNA	557
	Q	y	467	----APMKVGLBETAVERN-----DLISPPS-----GAPLISALGDLG-----	498
	D	b	558	NTTFTAGDYDIYPRKSVNGGFDDLGNANFYVPDGGAGITTPLKSPGVYTATEIYYVGNDIVT	617
	Q	y	499	-----GETHADVT-AVDSTIFTSPVFTTG-----TSVDYAELNP	531
	D	b	618	CQNQIFDGVFYGLOPVLNLTSAGSVIFTNNILANGSLQVOTAQNVISQPSSLSAIVEIAS	677
	Q	y	532	SITIIRAGSDPSQPNDRHVARF---STDGKNWFQSEPGCVTTG-----CTVA	577
	D	b	678	DVLINLAG-----QNVSFNGNTRGNVDIOAL--GNISTGSIVTSFPFGNAGNVI	725
	Q	y	578	ASADG---SRFWAPGDGPQPVVYAVVFGNSWA-----ASQGV	612
	D	b	726	LNAGGTLTCYIETSGTNCGDVITSSGNTSTAVIDTRGFDGLSIDSLGGAVSIESKGD	785
	Q	y	613	PANAQRISRVRNKFTVALSNCFYRSTDCGVTTPQVPAAGLPSSGAVGMFHAVPGKE--	670
	D	b	786	ITTAFTIDTCAYSIESFNEGTCGNVFLTAGDSITTVNYIFTAKNGG--DIPFOAGESIEII	843
	Q	y	671	-----GDLWLAA-----SSGLYHSHTNGGS-----SWSALTGV	697
	D	b	844	DYLNTVYSOTSGDBVVYEAPLDISIGSYIYTGCGGEPCGNVFLQAGDDITTYIDTSAANG	903
	Q	y	698	SSAVNVG-----FKKSAPGSSYPAVFVVGT	722
	D	b	904	DIFIQGSGTEVCYLTKGYEGRGDYYVETGRYFRACDGLLEECPFSVYTAGLVG-	962
	Q	y	723	IGGVTAQRSDDCGTTWVLINDQHQYGNWGQAITGDHANLRRYICTNCR--GIYVYGD	780
	D	b	963	-----GSVVI-----QFGGSEPFIIGNPI-----TNGTIGAISSGDD	994
	Q	y	781	GGAPEGSP-----SPVSFSASSLSPSPSSSPSPSPSPSPSPSSPSSPS	825
	D	b	995	NTPVIGTPIFDFTLLDNITITTEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPE	1054
	Q	y	826	PSPPSP	868
	D	b	1055	PE	1097
			RESULT 3		
			US-09-119-507B-112		
			; Sequence 112, Application US/09119507B		
			; Patent No. 6548642		
			; GENERAL INFORMATION:		
			; APPLICANT: Kieliszewski, Marcia J.		
			; TITLE OF INVENTION: No. 6548642el Synthetic Genes for Plant Gums		
			; FILE REFERENCE: OHU-03417		
			; CURRENT APPLICATION NUMBER: US/09/119,507B		
			; CURRENT FILING DATE: 1998-07-20		
			; NUMBER OF SEQ ID NOS: 118		
			; SOFTWARE: PatentIn Ver. 2.0		
			; SEQ ID NO 112		
			; LENGTH: 96		
			; TYPE: PRT		
			; ORGANISM: Artificial Sequence		
			; FEATURE:		
			; OTHER INFORMATION: Description of Artificial Sequence: Synthetic		
			US-09-119-507B-112		
			Query Match	5.6%; Score 288; DB 4; Length 96;	
			Best Local Similarity	77.0%; Pred. No. 1.le-11;	
			Matches	57; Conservative 4; Mismatches 13; Indels 0; Gaps 0;	
	Q	y	791	SVSPSASP	850
	D	b	20	SLSLAQTTRASPSFPSFPSFPSFPSFPSFPSFPSFPSFPSFPSFPSFPSFPSFPSFPS	79
	Q	y	851	SPSSSPSSGPSPTP	864

US-09-134-001C-4463
; Sequence 4463, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4463
; LENGTH: 2137
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4463

Query Match 4.6%; Score 234; DB 4; Length 2137;
Best Local Similarity 22.0%; Pred. No. 1.9e-06;
Matches 214; Conservative 129; Mismatches 404; Indels 226; Gaps 41;

QY 59 IGGGFGVDGIVNEGAPGILYVRTDIGGMYRWDAAAGRWI-----PLLDWYGNWNGY 111
DB 435 VGGGAP-----GA----FVSTDRNGMATTEASSAAKLNVQPTDNSQDFV----IDY 479
QY 112 NG-----VVSIAADPINTNKWAAVGMYNWSDPNDGAILRSSDQATWQITPLPKLGG 166
DB 480 NGDTKVMVTYAGQTTRN-----LTDWIKNSG-----GTTFSL-MTASTGG 521
QY 167 --NMPGRGGERLAVDPNDNIIYFCAPSCKGL-----197
DB 522 AKNLOQVQGTFEYTESAVAKRYVDANTGKOIIPKTIAGEVDAVNDKQNLNLSNG 581
QY 198 --WRSTDS--GATWSQMTNPPDVGTVIANPTDTGYQ--SDIQGVWVAPDKSSSLGQAS 252
DB 582 YSVSTDALQNSYSETSGTPTL--KLWSSQTVIYKFDVQV--PQISVDQSQREVGKTI 638
QY 253 KTFVGVADPNPVPVRSRGGATWQVPGAPTGFIPHKGVDPVNVHLVIATNTGPGYD 312
DB 639 NPITITTTDNSKDLVTT-----IVTGLPSGL-----SFDQTTWTI-----TGPSE 679
QY 313 -----GSSGVWKFVSTSGTWTTRISPV-----PSTDTANDYFG 345
DB 680 VGTTVTVNTDAGTGNVTSKQFTITTIQDTISPVVNVTPSQASEVFTPIINPITITATDMSG 739
QY 346 Y-----SGLTIDRQHPNTIMVATOISWPDITIFRSTP--GGATWTRI--WDWT-- 390
DB 740 KVTHVTGLPQKLFDASTNSIVGPTQIG--TNTITESTDASGNKTKTKINYEVRN 797
QY 391 --SYPNRSLRYV-----LDISAEPLWTFGVQPNPVPSPKLGWMDMAIDPFNSDRMLYGT 445
DB 798 SASDSTSTSVNSVSTSIKNSLSDSVKASQSLSTSKS--LSESLASSTNSSTSIQASE 855
QY 446 GAT-----LVATNDLTKWDSGGQTHIAPMKVGLBETAVNDLISPPSGLISALGDLGGFT 501
DB 856 SASSTKQLESASTSDSASE-----SARKSESTSKSTLSSESTSTSVSDSASVSTSES 910
QY 502 HADTVAVPSTIFTSVPVFTTGTSTVDYAEPLNPSIIVRAGSPDPSPQPNDRHVAESTDGGKNW 561
DB 911 ASTSTSVSGSTSTSIKNSLSTST-----SDSASIKASESASTSKLSESVSTST----- 959
QY 562 FQSGEPGVTTGGTVAASADGSRFVWAPDGPQVAVYVAVFGNSWAASQCVPANAIQRSD 621
DB 960 ----SDSASTSTSVSDNSASTSL-----SKSTSTSVSDSTSTSTSDSASTST---SE 1005
QY 622 RVNPKTFYALNGTFFRSTDDGGVTFQVVAAGLPSSGAVGMVHFVPGKSGDLWLAASGL 681
DB 1006 SESDSASTSLSESTSTSVSDSTSTSTSDSASMSASESES-----NSKSTSLSESTSTSL 1059

682 YHSTNGSSMSAITGVSSAVNVGFGKAPGSSYPAVFVVGTIGGVTGAYRSDDCGTTWVL 741
1060 SGTSTASTSDASTSTSESDSTSTSLSESTST-----SLSGSTASTSDASTS-TS 1112
QY 742 INDDQHOYGNWGOAI-----TGDHANLRRVYIGTNGRGI--VYGDIGCAPSGSPS 789
DB 1113 ESUSTSESTSLSESLSTSVSDSTASTSESTSESTSESTSESTSESTSESTSLSDSTSL 1172
QY 790 PSVPSASPLSPSS 849
DB 1173 TSTSDASTSTSESDSTSTSLSESTSTSLSDSTSTSTSESTSESTSESTSESTSL 1232
QY 850 PSPSSSPSS 909
DB 1233 ESTSTSVSDSTASTSDS--ASTSTSV-----SDSEASTSISESL-----STSVSDSTST 1281
QY 910 VRYWFTRDGGSST 922
DB 1282 ----STSDSASTS 1290

RESULT 8
US-09-033-537A-1
; Sequence 1, Application US/09033537A
; Patent No. 5958083
; GENERAL INFORMATION:
; APPLICANT: Onishi, Masahiro
; APPLICANT: Fich, Merete
; APPLICANT: Toft, Annette Hanne
; APPLICANT: Sh lein, Martin
; TITLE OF INVENTION: Prevention Of Back-Staining
; TITLE OF INVENTION: In Stone Washing
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5958083o No. 5958083disk of No. 5958083th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA: US/09/033,537A
; APPLICATION NUMBER: 0993/95
; FILING DATE: 08-SEP-1995
; APPLICATION NUMBER: PCT/DK96/00364
; FILING DATE: 03-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Green, Reza
; REGISTRATION NUMBER: 38,475
; REFERENCE/DOCKET NUMBER: 4492.204-US
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 551 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-033-537A-1

Query Match 4.4%; Score 225.5; DB 2; Length 551;
Best Local Similarity 23.8%; Pred. No. 1.1e-06;
Matches 114; Conservative 72; Mismatches 166; Indels 127; Gaps 22;

41 LMSRYSMLDQIKS--LGNTIRLPSDDILKPGTWPNSINFYQMNQDLQGTSLQVMD 98
481 LISPSPGAPLISALGDLG-----GFTHADVTAPSTIFTSPVFTTGTSDVYABL----- 529
99 KI-----VAYAGIQLRIILDRHPDCSGQALWYTSVSEATWISDLQALAQRYKG 150
530 NPSIIVRAGSFPDSSQPNDRHVAFSTGGKXWFGSGEPG-----VTTGGTVAA 578
151 NETVW-----GFDLHNEPHDPACWCGDPSIDWRLAAERAGNAVLNPNLLIEVEG--VQ 204
579 SADGSRFVWAPGDPQPVVAVGFGHSAASQ-----GVP-----ANQIRSDRVNPKTFY 629
205 SYNGDSYWRG-----GNLQAGAGYPPVVLNVENRLVYSADHYATSYPQTFW 250
630 ALSNGTFYRSTDCGVTFPQVAAGLPSSGAVGVMFHAPVPGKGDMLAASGLVHSTNGGS 689
251 --SDPTF-----PNV-----MPGLWNKW-----GYLFPNQNIAP 277
690 SMSAITG--VSSAVNVGFGKSAFSGSYPAVVVGTIGGVTGAYRSD-----DCG 736
278 VMLGEGFTLQSTTDQTLKTLVQLRP-----TAQXGADSFQWTFWSNPDG 326
737 TTWVLINDOQXGNGKQALTGHDHANLRVYIGTNGRGIVYGDIGGAPSGSPSPSPSA 796
327 DTGGILKDD-----WQVTDVTKDGYLAPI-----KSSIFDPVGA-----SA 362
797 SP 856
363 SP--SQSP 418
857 SSSP 915
419 ASGARCTAS-----YQVNSDWGNGFT-----VTVAVTNSGSAVKTKTWSVTF- 461
916 RUGGSSTLVNCDWAA 931
462 --GGNQITNS--WNA 473

RESULT 13
US-08-529-055-54
; Sequence 54, Application US/08529055
; Patent No. 6592876
; GENERAL INFORMATION:
; APPLICANT: Briles, David E.
; APPLICANT: McDaniel, Larry S.
; APPLICANT: Swiatlo, Edwin
; APPLICANT: Yother, Janet
; APPLICANT: Brooks-Walter, Alexis
; TITLE OF INVENTION: Pneumococcal Genes, Portions
; TITLE OF INVENTION: Thereof, Expression Products
; TITLE OF INVENTION: Therefrom, and Uses of Such Genes,
; TITLE OF INVENTION: Portions and Products
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford, P.C.
; STREET: 530 Fifth Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/529,055
; FILING DATE: 15-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Frommer, William S.

REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454312-2400
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 840-3333
TELEFAX: (212) 840-0712
INFORMATION FOR SEQ ID NO: 54:
SEQUENCE CHARACTERISTICS:
LENGTH: 206 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-529-055-54
Query Match 4.1%; Score 208.5; DB 4; Length 206;
Best Local Similarity 35.8%; Pred. No. 3.9e-06;
Matches 29; Conservative 31; Mismatches 20; Indels 1; Gaps 1;
Qy 784 PGSP 843
Db 101 PAPAPAPAPAPTPE-AP 159
Qy 844 PSPSSSP 864
Db 160 PAP 180
RESULT 14
US-09-198-956-10
; Sequence 10, Application US/09198956
; Patent No. 6165769
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels Erik K.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Schnorr, Kirk
; TITLE OF INVENTION: Pectin Degrading Enzymes From Bacillus
; FILE REFERENCE: 5377-200-US
; CURRENT APPLICATION NUMBER: US/09/198,956
; CURRENT FILING DATE: 1998-11-24
; EARLIER APPLICATION NUMBER: 1344/97
; EARLIER FILING DATE: 1997-11-24
; EARLIER APPLICATION NUMBER: 60/067,240
; EARLIER FILING DATE: 1997-12-02
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Bacillus licheniformis
US-09-198-956-10
Query Match 4.1%; Score 208.5; DB 3; Length 493;
Best Local Similarity 23.5%; Pred. No. 1.2e-05;
Matches 103; Conservative 62; Mismatches 149; Indels 125; Gaps 19;
Qy 568 GGVTTGGTVAASAD-----GSRFVWAPGDPQPVVAVGFG-NSWAASQGVPAQAQIRSDR 622
Db 63 GTTITNTSASKIDKDVSNVSVSGTGKE----LKGIGIKWRANNIIRNLKTHE-- 116
Qy 623 VNPKTFFYALSNGTFYRSTDCGVTFPQVAAGLPSSGAVGVMFHAPVPGKGDMLAASGLY 682
Db 117 -----VASG--DKDAIG-----IEGFSKNINWDHNE-LY 142
Qy 683 HSTNGGSSW-----SAITGVSSAVNVGFGKSAFSGSYPAVVVGTIGGTVGAYR 731
Db 143 HSLNVKDYDGLFDVKDEAEVITFSWYVHDGWSMLMGSS-----D 185
Qy 732 SDCGTTWVLND-----DQHYGNWQAITGDHANLR---RVYIGTNGR 773
Db 186 SDNYNRITTFHHNWFENLNSRVPSFRFGEGHTYNNYFNKIIDSGINSRMGARIRIENN-- 243

QY 774 GIVYDGGAPSGSPSPVS-PSASPS---LSPSP---SPSSSPSPSPSPSPSPSPSP 826
Db 244 -----LFEAKDPIVSWYSSPGYKWSNKNKFNNSRGSMPTTSTTTTYPYSYSLD 294
QY 827 SP 886
Db 295 NVDNVSIVKQ---NAGVGKIQRPPPTPTSPSPS---ANTPVSGNLKVEFYNSPDDTT 348
QY 887 NQIKPGLOVNTGSSVDLSVTVRYWFTRDGSSSTLVYNCDAWAAI-----GCGNI 937
Db 349 NSINPQFKVTNTGSSAIDLSKLTLYYYTVDGQKDTFW-CDHAAIIGSNGSYNGITSNV 407
QY 938 RASFGSVNPATPTADTYLQ 956
Db 408 KGTfVMSSTNNADTYLE 426

RESULT 15

US-09-198-955A-12
; Sequence 12, Application US/09198955A
; Patent No. 6187580
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schuelein, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. 6187580el Pectate Lyases
; FILE REFERENCE: 5378.200-US
; CURRENT APPLICATION NUMBER: US/09/198, 955A
; CURRENT FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 09/073,684
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217
; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Clostridium thermocellum
US-09-198-955A-12

Query Match 4.1%; Score 208.5; DB 3; Length 493;
Best Local Similarity 23.5%; Pred. No. 1.2e-05;
Matches 103; Conservative 62; Mismatches 149; Indels 125; Gaps 19;
QY 568 GGVTTGGTVAASAD---GSRFVWAPDGPQPVVAVGFG-NSWAAASQGVPAANAQIRSDR 622
Db 63 GTTTSNTSASKIDVKDVSNSIVSGTGKE-----LKGIGIKIWRANNIIRNLKHE-- 116
QY 623 VNPKTFYALNSGTFRSTGGVTFQPVAAAGLPSGAGVGMFHAVPGKEGDLWLAASSGLY 682
Db 117 -----VASG--DKDAIG-----IEGPKNIWVDHNE-LY 142
QY 683 HSTNGSSW-----SAITGVSSAVNVGFGKAPGSSYPVAVFVGTGIGVYGAYR 731
Db 143 HSLNVDKDYDGLDFVKRDAEYITFSWNYVHDGKSMLMGSS-----D 185
QY 732 SDDCGTTWLIND-----DQHQYGNWQAITGDHANLR---RVYIGTNGR 773

Db 186 SDTNRRTITTFHHNFENLSRVSPFRFGEGHYNNYFNKLIIDSGINSRMGARIRIENN-- 243
QY 774 GIVYDGGAPSGSPSPVS-PSASPS---LSPSP---SPSSSPSPSPSPSPSPSPSPSP 826
Db 244 -----LFEAKDPIVSWYSSPGYKWSNKNKFNNSRGSMPTTSTTTTYPYSYSLD 294
QY 827 SP 886
Db 295 NVDNVSIVKQ---NAGVGKIQRPPPTPTSPSPS---ANTPVSGNLKVEFYNSPDDTT 348
QY 887 NQIKPGLOVNTGSSVDLSVTVRYWFTRDGSSSTLVYNCDAWAAI-----GCGNI 937
Db 349 NSINPQFKVTNTGSSAIDLSKLTLYYYTVDGQKDTFW-CDHAAIIGSNGSYNGITSNV 407
QY 938 RASFGSVNPATPTADTYLQ 956
Db 408 KGTfVMSSTNNADTYLE 426

Search completed: May 14, 2004, 09:19:15
Job time : 27 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 14, 2004, 09:18:04 ; Search time 54 Seconds
(without alignments)
4931.411 Million cell updates/sec

Title: US-09-917-376-1
Perfect score: 5135
Sequence: 1 MDRSENIRLTWRSRLVSL.....RASFGSVNPATPTADTYLQX 957

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1145568 seqs, 278261457 residues
Total number of hits satisfying chosen parameters: 1145568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpa/US07_PUBCOMB.pap.*
- 2: /cgn2_6/ptodata/1/pubpa/PCT_NEW_PUB.pap.*
- 3: /cgn2_6/ptodata/1/pubpa/US06_NEW_PUB.pap.*
- 4: /cgn2_6/ptodata/1/pubpa/US06_PUBCOMB.pap.*
- 5: /cgn2_6/ptodata/1/pubpa/US07_NEW_PUB.pap.*
- 6: /cgn2_6/ptodata/1/pubpa/PCTUS_PUBCOMB.pap.*
- 7: /cgn2_6/ptodata/1/pubpa/US08_NEW_PUB.pap.*
- 8: /cgn2_6/ptodata/1/pubpa/US08_PUBCOMB.pap.*
- 9: /cgn2_6/ptodata/1/pubpa/US09A_PUBCOMB.pap.*
- 10: /cgn2_6/ptodata/1/pubpa/US09B_PUBCOMB.pap.*
- 11: /cgn2_6/ptodata/1/pubpa/US09C_PUBCOMB.pap.*
- 12: /cgn2_6/ptodata/1/pubpa/US09_NEW_PUB.pap.*
- 13: /cgn2_6/ptodata/1/pubpa/US10A_PUBCOMB.pap.*
- 14: /cgn2_6/ptodata/1/pubpa/US10B_PUBCOMB.pap.*
- 15: /cgn2_6/ptodata/1/pubpa/US10C_PUBCOMB.pap.*
- 16: /cgn2_6/ptodata/1/pubpa/US10_NEW_PUB.pap.*
- 17: /cgn2_6/ptodata/1/pubpa/US60_NEW_PUB.pap.*
- 18: /cgn2_6/ptodata/1/pubpa/US60_PUBCOMB.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5134	100.0	957	12	US-09-917-376-1 Sequence 1, Appli
2	5134	100.0	957	14	US-10-155-400-1 Sequence 1, Appli
3	4036	78.6	740	12	US-09-917-376-3 Sequence 3, Appli
4	4036	78.6	740	12	US-09-917-376-6 Sequence 6, Appli
5	4036	78.6	740	14	US-10-155-400-3 Sequence 3, Appli
6	4036	78.6	740	14	US-10-155-400-6 Sequence 6, Appli
7	2478	48.3	882	14	US-10-156-761-9395 Sequence 9395, Ap
8	1680	32.7	726	12	US-09-917-376-7 Sequence 7, Appli
9	1680	32.7	726	14	US-10-155-400-7 Sequence 7, Appli
10	1625.5	31.7	838	12	US-10-420-191-2 Sequence 2, Appli
11	1604	31.2	818	14	US-10-026-994-2 Sequence 2, Appli
12	1442	28.1	739	14	US-10-156-761-10111 Sequence 10111, A
13	1125	21.9	812	12	US-10-395-241-12 Sequence 12, Appl
14	1117	21.8	789	12	US-10-395-241-14 Sequence 14, Appl
15	1114	21.7	826	12	US-10-395-241-18 Sequence 18, Appl

16	848	16.5	555	10	US-09-927-827-47	Sequence 47, Appl
17	823	16.0	1228	10	US-09-917-384-1	Sequence 1, Appli
18	823	16.0	1228	10	US-09-917-383-1	Sequence 1, Appli
19	702	13.7	762	10	US-09-917-378-1	Sequence 1, Appli
20	511.5	10.0	1043	10	US-09-917-384-6	Sequence 6, Appli
21	511.5	10.0	1043	10	US-09-917-383-6	Sequence 6, Appli
22	462	9.0	88	12	US-09-917-376-5	Sequence 5, Appli
23	462	9.0	88	14	US-10-155-400-5	Sequence 5, Appli
24	462	9.0	89	12	US-09-917-376-4	Sequence 4, Appli
25	462	9.0	89	14	US-10-155-400-4	Sequence 4, Appli
26	462	9.0	154	10	US-09-917-378-4	Sequence 4, Appli
27	459	8.9	150	10	US-09-917-384-5	Sequence 5, Appli
28	459	8.9	150	10	US-09-917-383-5	Sequence 5, Appli
29	336.5	6.6	741	14	US-10-156-761-8100	Sequence 8100, Ap
30	288	5.6	96	15	US-10-437-708-236	Sequence 236, App
31	288	5.6	96	15	US-10-395-402-112	Sequence 112, App
32	262	5.1	2468	12	US-10-282-122A-66335	Sequence 66335, A
33	262	5.1	2468	14	US-10-246-330-4	Sequence 4, Appli
34	261.5	5.1	599	10	US-09-955-555A-29	Sequence 29, Appl
35	253	4.9	2435	12	US-10-282-122A-47453	Sequence 47453, A
36	252.5	4.9	2117	15	US-10-801-63	Sequence 63, Appl
37	243.5	4.7	1049	12	US-10-282-122A-49900	Sequence 49900, A
38	243	4.7	1259	14	US-10-260-715-8	Sequence 8, Appli
39	236	4.6	800	14	US-10-029-386-32198	Sequence 32198, A
40	232.5	4.5	1016	12	US-10-282-122A-69491	Sequence 69491, A
41	227.5	4.4	2271	12	US-10-282-122A-43924	Sequence 43924, A
42	226.5	4.4	1831	12	US-10-282-122A-71033	Sequence 71033, A
43	221.5	4.3	258	15	US-10-104-047-3034	Sequence 3034, Ap
44	221	4.3	2283	14	US-10-172-502-4	Sequence 4, Appli
45	219	4.3	1236	10	US-09-769-787-109	Sequence 109, App

ALIGNMENTS

RESULT 1
US-09-917-376-1
; Sequence 1, Application US/09917376
; Publication No. US20040038334A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS
; FILE REFERENCE: 40197.4US01
; CURRENT APPLICATION NUMBER: US/09/917,376
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 957
; TYPE: PRT
; ORGANISM: Acidothermus cellulosilyticus
; FEATURES:
; NAME/KEY: MOD_RES
; LOCATION: (957)
; OTHER INFORMATION: Any amino acid
US-09-917-376-1

Query Match 100.0%; Score 5134; DB 12; Length 957;
Best Local Similarity 100.0%; Pred. No. le-290;
Matches 956; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MDRSENIRLTWRSRLVSLAATASFAVAALGVLPPIAITASPAHAATQPTWNSVAIG 60
Db	1	MDRSENIRLTWRSRLVSLAATASFAVAALGVLPPIAITASPAHAATQPTWNSVAIG 60
Qy	61	GGGFVDGIVFNEGAGPILYVRTDIGMYRWDAANGRWIPLLDWGVGNWNGVWVSIAD 120
Db	61	GGGFVDGIVFNEGAGPILYVRTDIGMYRWDAANGRWIPLLDWGVGNWNGVWVSIAD 120

121	Qy	PINTNKVAAVAGMYTNSWDNDGAILRSSDQAGTWQIITPLPFKLGGNNPGRGMGRRLAVD	180
121	Db	PINTNKVAAVAGMYTNSWDNDGAILRSSDQAGTWQIITPLPFKLGGNNPGRGMGRRLAVD	180
181	Qy	PNNDNILYFGAPSGKGLWRSTDGATWSQMTNFPDVGTYIANPTDTTGYQSDIQGVWVA	240
181	Db	PNNDNILYFGAPSGKGLWRSTDGATWSQMTNFPDVGTYIANPTDTTGYQSDIQGVWVA	240
241	Qy	FDKSSSSILGQASKTIIFGVADPNPNPFWSRDGGATWQAVPGAGTGFIPHKGVDFPVNHVL	300
241	Db	FDKSSSSILGQASKTIIFGVADPNPNPFWSRDGGATWQAVPGAGTGFIPHKGVDFPVNHVL	300
301	Qy	YIATSNYGGPYDGSDDVWKSVSFTGWTTRISPVESDTDTANDYFGYSGLITIDROHPNTIM	360
301	Db	YIATSNYGGPYDGSDDVWKSVSFTGWTTRISPVESDTDTANDYFGYSGLITIDROHPNTIM	360
361	Qy	VATQISWNPDTIIFRSTDGGATWRIWDWTSYPNRSRYLVLDISAEPLWLTFCGQPNPVP	420
361	Db	VATQISWNPDTIIFRSTDGGATWRIWDWTSYPNRSRYLVLDISAEPLWLTFCGQPNPVP	420
421	Qy	SPKLGWMDDEMAIDPFNSDRMLYGTGATLYATNDLTKWDSGGQIHIAPMVKGLEETAVND	480
421	Db	SPKLGWMDDEMAIDPFNSDRMLYGTGATLYATNDLTKWDSGGQIHIAPMVKGLEETAVND	480
481	Qy	LI SPPSGAPLISALGDLGGFTHADVAVPSTIFTSPVFTTGTSDVDYAEINLPSIIVRAGSF	540
481	Db	LI SPPSGAPLISALGDLGGFTHADVAVPSTIFTSPVFTTGTSDVDYAEINLPSIIVRAGSF	540
541	Qy	DPSSQPNDRHVAFSTDDGKNWFOQSEPGCVTTGGTVAAASADGSRFVWAPGDPGPVYAV	600
541	Db	DPSSQPNDRHVAFSTDDGKNWFOQSEPGCVTTGGTVAAASADGSRFVWAPGDPGPVYAV	600
601	Qy	GFNSWAAASQGVPAANAQIRSDRVNPKTPYALNSGHTFYRSTDGGVTFPQVAAAGLPSGAGV	660
601	Db	GFNSWAAASQGVPAANAQIRSDRVNPKTPYALNSGHTFYRSTDGGVTFPQVAAAGLPSGAGV	660
661	Qy	VMFHVAPGKEGDLMLAASSGLYHSTNGGSSNSAITGVSSAVNVNFGKAPGSSYPAFVW	720
661	Db	VMFHVAPGKEGDLMLAASSGLYHSTNGGSSNSAITGVSSAVNVNFGKAPGSSYPAFVW	720
721	Qy	GTIGGVTVGAYRSDDCGTTTWVLINDDQHQYGNWQQAITGDHANLRVYIGTNGRGIVYGD	780
721	Db	GTIGGVTVGAYRSDDCGTTTWVLINDDQHQYGNWQQAITGDHANLRVYIGTNGRGIVYGD	780
781	Qy	GGAPSGSPSPSVSPASPSLSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP	840
781	Db	GGAPSGSPSPSVSPASPSLSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP	840
841	Qy	SASPSPSSSP	900
841	Db	SASPSPSSSP	900
901	Qy	SSVDLSTVTVRYWFTTRDGGSTLVYNCDDWAAIGCCGNIRASFGSVNPAFTADTYLQ	956
901	Db	SSVDLSTVTVRYWFTTRDGGSTLVYNCDDWAAIGCCGNIRASFGSVNPAFTADTYLQ	956

RESULT 2

[illegible]

Qy 901 SSVLDSTVTVRYWFTTRDGGSTLVYNCDMAAIGCGNIRASFGSVNRPATPTADTYLQ 956
|
Db 901 SSVLDSTVTVRYWFTTRDGGSTLVYNCDMAAIGCGNIRASFGSVNRPATPTADTYLQ 956

RESULT 3

US-09-917-376-3
; Sequence 3, Application US/09917376
; Publication No. US20040038334A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOYTICUS
; FILE REFERENCE: 40197.4US01
; CURRENT APPLICATION NUMBER: US/09/917,376
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 740
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:
; OTHER INFORMATION: Catalytic domain GH74

US-09-917-376-3

Query Match 78.6%; Score 4036; DB 12; Length 740;
Best Local Similarity 100.0%; Pred. No. 6.9e-227;
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 47 ATTQPYTWSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 106
Db 1 ATTQPYTWSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 60
Qy 107 NNWYNGVWSIAADPINTNKVAAVGMVYNSWDPNDGAILRSSDQATWQITPLPKLGG 166
Db 61 NNWYNGVWSIAADPINTNKVAAVGMVYNSWDPNDGAILRSSDQATWQITPLPKLGG 120
Qy 167 NMPGRGGERLAVDPNNNDILYFCAPSGKGLWSTDSGATWSQMTNFPDVGTYIANPTDT 226
Db 121 NMPGRGGERLAVDPNNNDILYFCAPSGKGLWSTDSGATWSQMTNFPDVGTYIANPTDT 180
Qy 227 TGYQSDIQGVVWVAFDKSSSLGQASKTIIFGVADPNPNPVFWSRDGGATWQAVPGAPTGF 286
Db 181 TGYQSDIQGVVWVAFDKSSSLGQASKTIIFGVADPNPNPVFWSRDGGATWQAVPGAPTGF 240
Qy 287 IPHKGVPDPVNHVLYIATSNITGGPYDGGSDGVWKFVSITSGTWTIRISVPSTDTANDYFGY 346
Db 241 IPHKGVPDPVNHVLYIATSNITGGPYDGGSDGVWKFVSITSGTWTIRISVPSTDTANDYFGY 300
Qy 347 SGLTIDRQHPTIMVATQISWMPDPTIIFRSTDDGGATWTRIDWMTSYPNRSLRYVLDISAE 406
Db 301 SGLTIDRQHPTIMVATQISWMPDPTIIFRSTDDGGATWTRIDWMTSYPNRSLRYVLDISAE 360
Qy 407 PWLTFGVQPNPPVPSPKLGMWDEAMAIIDPNSDRMLYGTGATLYATNDLTKWDSGGQIHI 466
Db 361 PWLTFGVQPNPPVPSPKLGMWDEAMAIIDPNSDRMLYGTGATLYATNDLTKWDSGGQIHI 420
Qy 467 APWKVGLTEETAANDLISPPSGAPLISALGDLGGFTHADVTAVPSTIPTSVPFTTGTSDY 526
Db 421 APWKVGLTEETAANDLISPPSGAPLISALGDLGGFTHADVTAVPSTIPTSVPFTTGTSDY 480
Qy 527 AELNPSIIVRAGSFDPSQPNDRHVAFTDGGKKNWFGSEPGVGTGTTGTTVAASADGSRFV 586
Db 481 AELNPSIIVRAGSFDPSQPNDRHVAFTDGGKKNWFGSEPGVGTGTTGTTVAASADGSRFV 540
Qy 587 WAPGDPQCPVVYAVGFGNSWAASQGVPPANAIIRSDRWNPKTFYALSNGTFFYRSTDDGVTF 646
Db 541 WAPGDPQCPVVYAVGFGNSWAASQGVPPANAIIRSDRWNPKTFYALSNGTFFYRSTDDGVTF 600

Qy 647 QPVAAGLPSSGAVGVMFHAFVPEKEDLMLAASSGLYHSTNGGSSWSAITGVSSAVNVGFG 706
Db 601 QPVAAGLPSSGAVGVMFHAFVPEKEDLMLAASSGLYHSTNGGSSWSAITGVSSAVNVGFG 660
Qy 707 KSAPGSSYPVAVVVGTTIGGVTGAYRSDDCGTTWVLINDDOHQYGNWQAITGDHANLRRV 766
Db 661 KSAPGSSYPVAVVVGTTIGGVTGAYRSDDCGTTWVLINDDOHQYGNWQAITGDHANLRRV 720
Qy 767 YIGTNGRGIVYGDIGGAPSG 786
Db 721 YIGTNGRGIVYGDIGGAPSG 740

RESULT 4

US-09-917-376-6
; Sequence 6, Application US/09917376
; Publication No. US20040038334A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOYTICUS
; FILE REFERENCE: 40197.4US01
; CURRENT APPLICATION NUMBER: US/09/917,376
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 740
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus

US-09-917-376-6

Query Match 78.6%; Score 4036; DB 12; Length 740;
Best Local Similarity 100.0%; Pred. No. 6.9e-227;
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 47 ATTQPYTWSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 106
Db 1 ATTQPYTWSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 60
Qy 107 NNWYNGVWSIAADPINTNKVAAVGMVYNSWDPNDGAILRSSDQATWQITPLPKLGG 166
Db 61 NNWYNGVWSIAADPINTNKVAAVGMVYNSWDPNDGAILRSSDQATWQITPLPKLGG 120
Qy 167 NMPGRGGERLAVDPNNNDILYFCAPSGKGLWSTDSGATWSQMTNFPDVGTYIANPTDT 226
Db 121 NMPGRGGERLAVDPNNNDILYFCAPSGKGLWSTDSGATWSQMTNFPDVGTYIANPTDT 180
Qy 227 TGYQSDIQGVVWVAFDKSSSLGQASKTIIFGVADPNPNPVFWSRDGGATWQAVPGAPTGF 286
Db 181 TGYQSDIQGVVWVAFDKSSSLGQASKTIIFGVADPNPNPVFWSRDGGATWQAVPGAPTGF 240
Qy 287 IPHKGVPDPVNHVLYIATSNITGGPYDGGSDGVWKFVSITSGTWTIRISVPSTDTANDYFGY 346
Db 241 IPHKGVPDPVNHVLYIATSNITGGPYDGGSDGVWKFVSITSGTWTIRISVPSTDTANDYFGY 300
Qy 347 SGLTIDRQHPTIMVATQISWMPDPTIIFRSTDDGGATWTRIDWMTSYPNRSLRYVLDISAE 406
Db 301 SGLTIDRQHPTIMVATQISWMPDPTIIFRSTDDGGATWTRIDWMTSYPNRSLRYVLDISAE 360
Qy 407 PWLTFGVQPNPPVPSPKLGMWDEAMAIIDPNSDRMLYGTGATLYATNDLTKWDSGGQIHI 466
Db 361 PWLTFGVQPNPPVPSPKLGMWDEAMAIIDPNSDRMLYGTGATLYATNDLTKWDSGGQIHI 420
Qy 467 APWKVGLTEETAANDLISPPSGAPLISALGDLGGFTHADVTAVPSTIPTSVPFTTGTSDY 526
Db 421 APWKVGLTEETAANDLISPPSGAPLISALGDLGGFTHADVTAVPSTIPTSVPFTTGTSDY 480
Qy 527 AELNPSIIVRAGSFDPSQPNDRHVAFTDGGKKNWFGSEPGVGTGTTGTTVAASADGSRFV 586

Db 481 AELNPSIIVRAGSFSSQPNDRHVAFTDGGKWFQSGSEPGVTTGGTVAASADGSRFV 540
QY 587 WAPGDPQPVVAVGFGNSWAASQGVPAQAIRSDRVNPKTFYALNGTFYRSTDDGVTF 646
Db 541 WAPGDPQPVVAVGFGNSWAASQGVPAQAIRSDRVNPKTFYALNGTFYRSTDDGVTF 600
QY 647 QPVAAGLPSSGAVGVMFHAPVPGKEGDLWLAASSGLYHSTNGSSWSAITGVSAVNVGFG 706
Db 601 QPVAAGLPSSGAVGVMFHAPVPGKEGDLWLAASSGLYHSTNGSSWSAITGVSAVNVGFG 660
QY 707 KSAFGSSYPVAVVGTIGGVTGAYRSDCGTTWVLINDDQHOYGNWQOAITGDHANLRV 766
Db 661 KSAFGSSYPVAVVGTIGGVTGAYRSDCGTTWVLINDDQHOYGNWQOAITGDHANLRV 720
QY 767 YIGTNGRGIVYGDIGGAPSG 786
Db 721 YIGTNGRGIVYGDIGGAPSG 740
RESULT 5
US-10-155-400-3
; Sequence 3, Application US/10155400
; Publication No. US20030108988A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS
; FILE REFERENCE: NREL 01-36A
; CURRENT APPLICATION NUMBER: US/10/155,400
; CURRENT FILING DATE: 2002-10-22
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 3
; LENGTH: 740
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:
; OTHER INFORMATION: Catalytic domain GH74
US-10-155-400-3
Query Match 78.6%; Score 4036; DB 14; Length 740;
Best Local Similarity 100.0%; Pred. No. 6.9e-227;
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 47 ATTQPYTWSNVAIGGGGFVDGIVFNEGAPGILYVRTDIGMYRWDANGRWIPLLDVWG 105
Db 1 ATTQPYTWSNVAIGGGGFVDGIVFNEGAPGILYVRTDIGMYRWDANGRWIPLLDVWG 60
QY 107 NNWYNGVVSIAADPINTNKWAAVGMVYNSWDPNDGAILRSSDQGTATWQITPLPFKLG 166
Db 61 NNWYNGVVSIAADPINTNKWAAVGMVYNSWDPNDGAILRSSDQGTATWQITPLPFKLG 120
QY 167 NMPGRGMRERLAVDPNNNDILYFGAPSGKGLWRSTDSGATWSQMTNPPDVGTYIANPTDT 226
Db 121 NMPGRGMRERLAVDPNNNDILYFGAPSGKGLWRSTDSGATWSQMTNPPDVGTYIANPTDT 180
QY 227 TGYQSDIQGVVWVAFDKSSSLGQASKTIFVGVADPNNPVFWSRDCGATWQAVPGATGF 286
Db 181 TGYQSDIQGVVWVAFDKSSSLGQASKTIFVGVADPNNPVFWSRDCGATWQAVPGATGF 240
QY 287 IPHKGVDPVNVHLYIATSNTPGPDGSGDVWKFVSTGWTTRISPVSTDTANDYFGY 346
Db 241 IPHKGVDPVNVHLYIATSNTPGPDGSGDVWKFVSTGWTTRISPVSTDTANDYFGY 300
QY 347 SGLTIDROHNTIMVATQISWEPDTIIFRSTDCGATWTRIDWTSYPNRSLRYVLDISAE 406
Db 301 SGLTIDROHNTIMVATQISWEPDTIIFRSTDCGATWTRIDWTSYPNRSLRYVLDISAE 360
QY 407 PWLTGFGVQPNPPVPSPKLGWDEAMAIDPFNSDRMLYGTGATLYATNDLTKWDSGGQIHI 466

Db 361 PWLTGFGVQPNPPVPSPKLGWDEAMAIDPFNSDRMLYGTGATLYATNDLTKWDSGGQIHI 420
QY 467 APWKGLIETEAVNDLISPPSGAPLISALGDLGGFTHADVAVPSTIFTSPVFTTGTSDY 526
Db 421 APWKGLIETEAVNDLISPPSGAPLISALGDLGGFTHADVAVPSTIFTSPVFTTGTSDY 480
QY 527 AELNPSIIVRAGSFSSQPNDRHVAFTDGGKWFQSGSEPGVTTGGTVAASADGSRFV 586
Db 481 AELNPSIIVRAGSFSSQPNDRHVAFTDGGKWFQSGSEPGVTTGGTVAASADGSRFV 540
QY 587 WAPGDPQPVVAVGFGNSWAASQGVPAQAIRSDRVNPKTFYALNGTFYRSTDDGVTF 646
Db 541 WAPGDPQPVVAVGFGNSWAASQGVPAQAIRSDRVNPKTFYALNGTFYRSTDDGVTF 600
QY 647 QPVAAGLPSSGAVGVMFHAPVPGKEGDLWLAASSGLYHSTNGSSWSAITGVSAVNVGFG 706
Db 601 QPVAAGLPSSGAVGVMFHAPVPGKEGDLWLAASSGLYHSTNGSSWSAITGVSAVNVGFG 660
QY 707 KSAFGSSYPVAVVGTIGGVTGAYRSDCGTTWVLINDDQHOYGNWQOAITGDHANLRV 766
Db 661 KSAFGSSYPVAVVGTIGGVTGAYRSDCGTTWVLINDDQHOYGNWQOAITGDHANLRV 720
QY 767 YIGTNGRGIVYGDIGGAPSG 786
Db 721 YIGTNGRGIVYGDIGGAPSG 740
RESULT 6
US-10-155-400-6
; Sequence 6, Application US/10155400
; Publication No. US20030108988A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS
; FILE REFERENCE: NREL 01-36A
; CURRENT APPLICATION NUMBER: US/10/155,400
; CURRENT FILING DATE: 2002-10-22
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 6
; LENGTH: 740
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
US-10-155-400-6
Query Match 78.6%; Score 4036; DB 14; Length 740;
Best Local Similarity 100.0%; Pred. No. 6.9e-227;
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 47 ATTQPYTWSNVAIGGGGFVDGIVFNEGAPGILYVRTDIGMYRWDANGRWIPLLDVWG 106
Db 1 ATTQPYTWSNVAIGGGGFVDGIVFNEGAPGILYVRTDIGMYRWDANGRWIPLLDVWG 60
QY 107 NNWYNGVVSIAADPINTNKWAAVGMVYNSWDPNDGAILRSSDQGTATWQITPLPFKLG 166
Db 61 NNWYNGVVSIAADPINTNKWAAVGMVYNSWDPNDGAILRSSDQGTATWQITPLPFKLG 120
QY 167 NMPGRGMRERLAVDPNNNDILYFGAPSGKGLWRSTDSGATWSQMTNPPDVGTYIANPTDT 226
Db 121 NMPGRGMRERLAVDPNNNDILYFGAPSGKGLWRSTDSGATWSQMTNPPDVGTYIANPTDT 180
QY 227 TGYQSDIQGVVWVAFDKSSSLGQASKTIFVGVADPNNPVFWSRDCGATWQAVPGATGF 286
Db 181 TGYQSDIQGVVWVAFDKSSSLGQASKTIFVGVADPNNPVFWSRDCGATWQAVPGATGF 240
QY 287 IPHKGVDPVNVHLYIATSNTPGPDGSGDVWKFVSTGWTTRISPVSTDTANDYFGY 346
Db 241 IPHKGVDPVNVHLYIATSNTPGPDGSGDVWKFVSTGWTTRISPVSTDTANDYFGY 300

QY 347 SGLTIDRQHPNTIMWATQISWPDPTIIFRSTDCGATWTRIDWNTSYNRSIRYVLDISAE 406
DB 301 SGLTIDRQHPNTIMWATQISWPDPTIIFRSTDCGATWTRIDWNTSYNRSIRYVLDISAE 360
QY 407 PWTFTGVQPNPPVPSKLGWMDRMAIDPNSDRMLYGTGATLYATNDLTKWDSGGQIHI 466
DB 361 PWTFTGVQPNPPVPSKLGWMDRMAIDPNSDRMLYGTGATLYATNDLTKWDSGGQIHI 420
QY 467 APVKGLIETAVNDLISPPSGAPLISALGDLGGFTHADVTAVPSTIIFTSVPVFTGTSDVY 526
DB 421 APVKGLIETAVNDLISPPSGAPLISALGDLGGFTHADVTAVPSTIIFTSVPVFTGTSDVY 480
QY 527 AELNPSIIVRAGSFDPSQPNDRHVAFSTDCGKNWFGSPPGGVTTGGTVAASADGRFV 586
DB 481 AELNPSIIVRAGSFDPSQPNDRHVAFSTDCGKNWFGSPPGGVTTGGTVAASADGRFV 540
QY 587 WAPGDCQPVVYAVFGNSWAASQGVPANAIQIRSDRVNPKTFYALSNGTYRSTDCGVTF 646
DB 541 WAPGDCQPVVYAVFGNSWAASQGVPANAIQIRSDRVNPKTFYALSNGTYRSTDCGVTF 600
QY 647 QPVAAGLPSSGAVGVWFMHAPVPGKEDLWLAASGLYHSTNGSSWSAITGVSSAVNVGFG 706
DB 601 QPVAAGLPSSGAVGVWFMHAPVPGKEDLWLAASGLYHSTNGSSWSAITGVSSAVNVGFG 660
QY 707 KSAPGSSYPVAVVGGTGGTGYRSDCGTWTWINDDQHQYGNWGOAITGDHANLRV 766
DB 661 KSAPGSSYPVAVVGGTGGTGYRSDCGTWTWINDDQHQYGNWGOAITGDHANLRV 720
QY 767 YIGTNGRGIVYDIGGAPSG 786
DB 721 YIGTNGRGIVYDIGGAPSG 740

RESULT 7

US-10-156-761-9395
; Sequence 9395, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HAITOKI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 9395
; LENGTH: 882
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-9395

Query Match 48.3%; Score 2478; DB 14; Length 882;
Best Local Similarity 50.4%; Pred. No. 4.8e-136;
Matches 473; Conservative 128; Mismatches 227; Indels 110; Gaps 14;

QY 23 TASFAVAAL--GVLPATITASPAHAATTQPYTWSNVAGGGFVDCIVFNECAPGLYV 80
DB 8 TAVLALAAGLPACTPPALAAATATIAADTYSWKARVDGGFVGVFVNRSEKILAYA 67
QY 81 RTDIGMYRWAAGRWIPLLDVGWNNWNGYVIAADPINTKNKVAAGMYTNSWDP 140
DB 68 RTDIGGAYRWAESSKTWPTLLDSVGWSDWCHTGVVSLASDSDVPNPKYAAVGYTNSWDP 127

QY 141 NDGAILRSSDOGATWQITPLPKLGGNMPGRGMRERLAVDPNNDNILYFGAPSGKGLWS 200
DB 128 GNGAVLRSRSDRGASQWKTDLFPKLGNNMPGRGMRERLAVDPNRNVLYLFGAPSGKGLWS 187
QY 201 TDSGATWSQMTNFPDVGTYIANPDTTTCYQSDIQGVWVAFDKSSSSSLGQASKTTFVGV 260
DB 188 TDSGASWSQVTFPFPVGGIYQDADTSGYASDNQGIWVTFDESTGPGSSRTTVVGV 247
QY 261 DPNNPFWMSRDGGATWQAVPGAPTGFIPHKGVDFPVNHVLYIATSNTPGYPDGSSGDVWK 320
DB 248 DKONSVYRSTDAAGTWSRLAQPTGHLAKGVLDAAAGCLYLAAYSDKGPYDGGKQLWR 307
QY 321 FSVTSGTWTIRISPVSTDTANDYFGYSLTIDRQHPNTIMWATQISWPDPTIIFRSTDCG 380
DB 308 YTTKTGTWNTISPAEADT---YTGFSGLTVDRQHPGTVMATAYSSWNPDTQLFRSTDCG 364
QY 381 ATWTRIDWNTSYNRSIRYVLDISAEPLITGVQPNPPVPSKLGWMDRMAIDPNSDR 440
DB 365 GTWTKAWDYTSYPSRSNRFTMDVSSPWLWGANPAPPEQTCLKGMWTESLEIDPFDSAR 424
QY 441 MLYGTGATLYATNDLTKWDSGGQIHIAPMVKLEBETAVNDLISPPS--GAPLISALGDLGG 499
DB 425 MMYGTGATVYGTNDLTKWDSGQFTIKPMARGLEBETAVNDLASPSGGQLFSALGDLGG 484
QY 500 FTHADVTAVPSTIIFTSVPVFTTGTSDVYAEALNPSIIVRAGSFDPSQPNDRHVAFSTDCG 559
DB 485 FRHTDLITVPSLMYTSNFTTSTSLDYAETDGTVVVRVGNLD--SGP---HVAFSTDCG 539
QY 560 NWFPGSPGGVTTGGTVAASADGRFVWAPDGPQPVVYAVFGNSWAASQGVPANAIQIR 619
DB 540 NWFAGADPSGVSGGTVAASADGRFVWSPAGTG--VQYTTGFGTWSASAGLPAGAI 597
QY 620 SDRVNPKEFALSNGTFYRSTDCGVTQPVAA--GLPSSGAVGVWFMHAPVPGKEDLWLAAS 678
DB 598 SDRVDPKTFYFGKSGRFVYSSDGGATFTASATGLPSGDS--VRPKALPFTKGDILWLAG 655
QY 679 S-----GLYHSTNGSSWSAITGVSSAVNVGFGKSAAGSSYPVAVVGGTGGTGYRSD 733
DB 656 ASDGAYGLWHSITDGAATKLAIVDQADTIGFKAATGASVQTLTYSAKIGGVGRIFRST 715
QY 734 DCGTTWVLIINDDQHQYGNWGOAITGDHANLRVYIGTNGRGIVYDGIAGPSGSPSPSV 793
DB 716 DKGASWTRVNDDAHQWGTGAAITGDPVRYGVVYVSTNGRGIVYDGTAGSSDCG----- 769
QY 794 PSASPSLSPPSP 853
DB 770 ----- 769
QY 854 SSPSSSP 913
DB 770 ---GTEPAPT-----GACTVYTRITNMQSGGFQ--ADVQLANTGSTAWDQWSIG---W 814
QY 914 FTRDGGSSTLVYNCDDWAAIGCG-----NIRASFGS 943
DB 815 SFGDQGVETQLMNASYAGAGSGVTAANLWNRVAAGS 852

RESULT 8

US-09-917-376-7
; Sequence 7, Application US/09917376
; Publication No. US20040038334A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL B.
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS
; FILE REFERENCE: 40197.4US01
; CURRENT APPLICATION NUMBER: US/09/917,376
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1

```

; SEQ ID NO 7
; LENGTH: 726
; TYPE: PRT
; ORGANISM: Aspergillus aculeatus
US-09-917-376-7

Query Match      32.7%; Score 1680; DB 12; Length 726;
Best Local Similarity 46.2%; Pred. No. 1.1e-89;
Matches 346; Conservative 113; Mismatches 258; Indels 32; Gaps 17;

QY 47 ATTQPTWNSVAI--GGGGFVDGIVNEGAGLIVRTDGGMYRWDAAANGRIPLLDWVG 105
DB 1 AASQAYTKNVVTTGGGGGFTPGIVFNPISAKGVARTDIGGAYRLN--SDTWTPLMDWVG 59
QY 106 ---WNNNGYNGVSVIAADPINTNKVAAVGMVYNSWDPNDGAILRSSDQATWQITPLPF 162
DB 60 NDTHWDM---GIDALATDPVDTDRVYVAVGMVYNEWDPNVGSILRSDDQGTWETKLPF 116
QY 163 KLGNNPGRGMRGLAVDPNNNDILYFGAPSGKGLWRSTDSGATWSQMTNPPDVGTIYAN 222
DB 117 KVGNNPGRGMRGLAVDPNKNISILYFGARSGHGLWKS TDYGATWSNVTSFTWTGYFQD 176
QY 223 PTDITTYGSDIQGVVWVAFPKSSSLGQASKTI FVGVDAPNNPVFWSRDCGATWQAVPGA 282
DB 177 SSST--YTSDPVGIAMVTFDSTSGSSGATPRIFVGVADAGKSVFKSEDAAGATWAWVSGE 234
QY 283 PT-GFIPHKGVDPVNVHLYIATNSGTGPGYDGSQDVWKFVSVTGTRISPPVSTDTAN 341
DB 235 PQYGFPLPHKGVLSPEEKTLYISYANGAGPYDGTNGTVHKYNTISGVWTDISP---TSLAS 291
QY 342 DYFGYGLTIDRQHPNTIMVATQISWMPDTIIFRSTDCGATWTRIDWTSYPNRSRLRYVL 401
DB 292 TTYGYGGLSVLDQVPGTLMVAALNCWPDDELIFRSTDSGATWSPIMWNGVPSINYYSY 351
QY 402 DISAEPWLTGFGVQPNP--PVSPKLGWMDMAIDPFNSDRMLYCTGATLYATNDLTKWDS 460
DB 352 DISNAPMIQDITSTDDQFPV---RVGMVMEALADPFDSNHWLYGTLTVYGGHDLTNWDS 408
QY 461 GGOIHIAPMVKGLEETAANDLISPPSGAPLISALGDLGGFTHADVTAVPSTIFTSPVFTT 520
DB 409 KHNVTVKSIAVGEIEMAVLGLITPPGPPALLSAVGDDGCFYHSDLDAAPNQAYHTPYGT 468
QY 521 GTSVDYAEALNPSIIVRAGSPDPSSQPNDRHVAFSTDGKKNWFOGSEPGVTTGGTVAASA 580
DB 469 TNGIDYAGNKPNSIVRSGASD--DYPT--LALSSNFGSTWYADYAASTGTGVALSA 523
QY 581 DGRFRFWAPGDPQPVVAVGFGNSWAASQGVAPANAQIRSDRVNPKTFYALNSGTFFRST 640
DB 524 DGDVTLLMSSTSGALVSKSQG---TLTAVSLPSGAVIASDKSDNTVYFGSAGAIYVSK 580
QY 641 DGGVTQFPVAAGLPSSGAVGMFHAVPGKEGDLWLAASSGLYHSTNGSGSSWSAI--TCVSS 699
DB 581 NTATSFKTIVS--LGSSTTVNAI--RAHPSIAGDVWASTDKGLMHSSTDYSGTFTQIGSGVTA 638
QY 700 AVNVGFGKSPAGSSYPVAVFVGTIGVGTGAYRSDCCGTTWVLINDDOHQYGNWQA--IT 757
DB 639 GWGFGFGKASSTGYSVVIYGFFTIDGAAGLKFSEDACTNMQVSDASHGFGS--GSANVNV 697
QY 758 GDHANLRVYGTNGRGIVYGDIGGAPSG 786
DB 698 GDLQTYGRVFRGHERPGHLLRQSQREPAG 726

```

RESULT 9

```

US-10-155-400-7
; Sequence 7, Application US/10155400
; Publication No. US2003010898A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS

```

```

; TITLE OF INVENTION: CELLULOXYLICUS
; FILE REFERENCE: NREL 01-36A
; CURRENT APPLICATION NUMBER: US/10/155.400
; CURRENT FILING DATE: 2002-10-22
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 726
; TYPE: PRT
; ORGANISM: Aspergillus aculeatus
US-10-155-400-7

```

```

Query Match      32.7%; Score 1680; DB 14; Length 726;
Best Local Similarity 46.2%; Pred. No. 1.1e-89;
Matches 346; Conservative 113; Mismatches 258; Indels 32; Gaps 17;

QY 47 ATTQPTWNSVAI--GGGGFVDGIVNEGAGLIVRTDGGMYRWDAAANGRIPLLDWVG 105
DB 1 AASQAYTKNVVTTGGGGGFTPGIVFNPISAKGVARTDIGGAYRLN--SDTWTPLMDWVG 59
QY 106 ---WNNNGYNGVSVIAADPINTNKVAAVGMVYNSWDPNDGAILRSSDQATWQITPLPF 162
DB 60 NDTHWDM---GIDALATDPVDTDRVYVAVGMVYNEWDPNVGSILRSDDQGTWETKLPF 116
QY 163 KLGNNPGRGMRGLAVDPNNNDILYFGAPSGKGLWRSTDSGATWSQMTNPPDVGTIYAN 222
DB 117 KVGNNPGRGMRGLAVDPNKNISILYFGARSGHGLWKS TDYGATWSNVTSFTWTGYFQD 176
QY 223 PTDITTYGSDIQGVVWVAFPKSSSLGQASKTI FVGVDAPNNPVFWSRDCGATWQAVPGA 282
DB 177 SSST--YTSDPVGIAMVTFDSTSGSSGATPRIFVGVADAGKSVFKSEDAAGATWAWVSGE 234
QY 283 PT-GFIPHKGVDPVNVHLYIATNSGTGPGYDGSQDVWKFVSVTGTRISPPVSTDTAN 341
DB 235 PQYGFPLPHKGVLSPEEKTLYISYANGAGPYDGTNGTVHKYNTISGVWTDISP---TSLAS 291
QY 342 DYFGYGLTIDRQHPNTIMVATQISWMPDTIIFRSTDCGATWTRIDWTSYPNRSRLRYVL 401
DB 292 TTYGYGGLSVLDQVPGTLMVAALNCWPDDELIFRSTDSGATWSPIMWNGVPSINYYSY 351
QY 402 DISAEPWLTGFGVQPNP--PVSPKLGWMDMAIDPFNSDRMLYCTGATLYATNDLTKWDS 460
DB 352 DISNAPMIQDITSTDDQFPV---RVGMVMEALADPFDSNHWLYGTLTVYGGHDLTNWDS 408
QY 461 GGOIHIAPMVKGLEETAANDLISPPSGAPLISALGDLGGFTHADVTAVPSTIFTSPVFTT 520
DB 409 KHNVTVKSIAVGEIEMAVLGLITPPGPPALLSAVGDDGCFYHSDLDAAPNQAYHTPYGT 468
QY 521 GTSVDYAEALNPSIIVRAGSPDPSSQPNDRHVAFSTDGKKNWFOGSEPGVTTGGTVAASA 580
DB 469 TNGIDYAGNKPNSIVRSGASD--DYPT--LALSSNFGSTWYADYAASTGTGVALSA 523
QY 581 DGRFRFWAPGDPQPVVAVGFGNSWAASQGVAPANAQIRSDRVNPKTFYALNSGTFFRST 640
DB 524 DGDVTLLMSSTSGALVSKSQG---TLTAVSLPSGAVIASDKSDNTVYFGSAGAIYVSK 580
QY 641 DGGVTQFPVAAGLPSSGAVGMFHAVPGKEGDLWLAASSGLYHSTNGSGSSWSAI--TCVSS 699
DB 581 NTATSFKTIVS--LGSSTTVNAI--RAHPSIAGDVWASTDKGLMHSSTDYSGTFTQIGSGVTA 638
QY 700 AVNVGFGKSPAGSSYPVAVFVGTIGVGTGAYRSDCCGTTWVLINDDOHQYGNWQA--IT 757
DB 639 GWGFGFGKASSTGYSVVIYGFFTIDGAAGLKFSEDACTNMQVSDASHGFGS--GSANVNV 697
QY 758 GDHANLRVYGTNGRGIVYGDIGGAPSG 786
DB 698 GDLQTYGRVFRGHERPGHLLRQSQREPAG 726

```

RESULT 10

```

US-10-420-191-2
; Sequence 2, Application US/10420191
; Publication No. US20040067569A1

```

GENERAL INFORMATION:
; APPLICANT: No. US20040067569Allozymes Biotech, Inc.
; APPLICANT: Rey, Michael W.
; APPLICANT: Zaretsky, Elizabeth J.
; APPLICANT: Haas, Jeffrey A.
; TITLE OF INVENTION: Polypeptides Having Xyloglucanase Activity And Nucleic Acids
; TITLE OF INVENTION: Encoding Same
; FILE REFERENCE: 10210.200-US
; CURRENT APPLICATION NUMBER: US/10/420,191
; CURRENT FILING DATE: 2003-04-18
; PRIOR APPLICATION NUMBER: US 60/373,987
; PRIOR FILING DATE: 2002-04-19
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 838
; TYPE: PRT
; ORGANISM: Trichoderma reesei
US-10-420-191-2

Query Match 31.7%; Score 1625.5; DB 12; Length 838;
Best Local Similarity 41.7%; Pred. No. 2e-86;
Matches 350; Conservative 131; Mismatches 296; Indels 63; Gaps 23;

QY 25 SFVAALGVLPIAITASPAAHAATTPYTWNSVAI-GGGGFVDGIVFNEGAPGILYVRTD 83
DB 4 SRVLALVIGAV-----IPAHAA-----FSWKNVKLGGGGFVPGIIFHPKTKGVAVARTD 53

QY 84 IGMRYRDAANGRWIPLLDWV-----GWNNGYGVVSIADPINTNKVWAAVGMVYNSWD 139
DB 54 IGGYLRIN-ADDSWTAVTGDIADNAGHNW---GIDAVALDPDQDKVYAAVGMVYNSWD 109

QY 140 PNDGAILRSSDCATQWITPLPKLGMPGRGWERLAVDPNNDLILYFGASGGKGLWR 199
DB 110 PSNGAILRSSDRGATWFTNLPPKVGMPGRGAGERLAVDPANSNIIFYGASGGKGLWK 169

QY 200 STDGATWMSQMTNFPDVGTYIANPTDTTQSDIQGVVWAFKSSSSLSQAQSKTIFVGV 259
DB 170 STDGVTFFKVSSTATGTVIPDPSDNGYNSDKQGLMWTFDSTSTTGATSRIFVGT 229

QY 260 ADP-NNPVFWSRGGATQWAVPGAPTGFIPHKGVFDPVNHVLIATSNTPGVDGSGDV 318
DB 230 ADNITASVYVSTNAGSTWAVPGQPKYFPHKAKLOPAERKALYLTSDGTGPDGTLGSV 289

QY 319 WKFSVSGTWTRISVPSTDTANDYFCYGLTIDRQHPNTIMWATQISWPDITIIIRSTD 378
DB 290 WRYDIAAGTWKDTITPVSGSL---YFGFGLGLDLQKPGTLVVASLNSWWPDQAFLPRSTD 346

QY 379 GGATWTRIMDWTSPNRSRLRYLDISAEPLWTFG-VQPNPPVPS---PKLGMWDEAMAI 433
DB 347 SGTTWSPINAWASYPTETYYISSTPKAPWKNFIDVTSESFSDGLIKELGHWIISLEI 406

QY 434 DPFNSRMLYGTGATLYATNDLTWDSGGQIHIAPMVKGLEETAANDLISPPSGAPLISA 493
DB 407 DPTDSNHWLYGTGWTIFGGHDLTNWDRHNSIQSLADGIEEFSVQDLASAPGSELLAA 466

QY 494 LGDLGGFTHA---DVTAVPSTIFTSVFTTGTSDVYAEALNPSIIVRAGSFPDSSQPNDRH 550
DB 467 VGDNDGFTFASRNDLGTSPQWATPWTATSTVYAGNSVKSVVRVGNVGTAGTQQ----- 521

QY 551 VAFSTDGKNWFGQSBPGGVTTGCTVAASADSRFVWAPGDPQPVVYAVFGFNSWAASQ 610
DB 522 VALSSDGGATWSDIYAADTSMNGGTVAISADGDTILWSTASSG---VQPSQFGSGFASVS 578

QY 611 GVPANAQIRSDRVNPKTFYALSNGTFYRSYTDGVTQPVAAAGLPSSGAVGVW---FHAVPG 668
DB 579 SLPAGAVIASDKKTNVFFYAGSGSTFVVKDTCGSSP---TRG-PKLGSACTIRDIAAHT 634

QY 669 KEGDLMLAASSGLYHSTNGGSSWAI-TGVSSAVNVGFGKAPGSSYPAVFVGTGTGGVT 727
DB 635 TAGTLYVSTDVGIFRSTDSGTTFQGVSTALTNTYQIALGVGS-GSNW-NLYAFGT--GPS 690

QY 728 GA--YRSDDCGTTWVLINDDQHQYGNWGOAITGDHANLRVYIGTNGRIVY--GDIGGA 783

DB 691 GARLYASGDSGASWTDIQSGQFGSIDSTKVAGSGSTAGQVTVGNRGRVFAQGTGGG 750
QY 784 PSGSPSPVSPSPASPSLSPPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPAS 843
DB 751 TGGT-----SSSTKQSSSTSSASSSTLRSSVSVVTRASTVTSRTSSAAGPTGS 801

RESULT 11
US-10-026-994-2
; Sequence 2, Application US/10026994
; Publication No. US20030113732A1
; GENERAL INFORMATION:
; APPLICANT: Dunn-Coleman, Nigel
; APPLICANT: Goedegebuur, Frits
; APPLICANT: Ward, Michael
; APPLICANT: Yao, Jian
; TITLE OF INVENTION: EGV1 Endoglucanase and Nucleic Acids
; TITLE OF INVENTION: Encoding the Same
; FILE REFERENCE: GC698
; CURRENT APPLICATION NUMBER: US/10/026,994
; CURRENT FILING DATE: 2002-04-30
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 818
; TYPE: PRT
; ORGANISM: Trichoderma reesei
US-10-026-994-2

Query Match 31.2%; Score 1604; DB 14; Length 818;
Best Local Similarity 41.8%; Pred. No. 3.5e-85;
Matches 340; Conservative 128; Mismatches 291; Indels 54; Gaps 21;

QY 52 YTSNVAI-GGGGFVDGIVFNEGAPGILYVRTDTCGMYRDAANGRWIPLLDWV---CW 106
DB 2 FSWKNVKLGGGGFVPGIIFHPKTKGVAYARTDIGLYRLN-ADDSWTAVTGDIADNAGW 60

QY 107 NNGCYGVVSIADPINTNKVWAAVGMVYNSWDPNNDGAILRSSDCATQWITPLPKLG 166
DB 61 HNW---GIDAVALDPDQDKVYAAVGMVYNSWDPSNGAILRSSDRGATWFTNLPPKVG 117

QY 167 NMPGRGGERLAVDPNNDLILYFGAPSGKGLWRSTDGATWMSQMTNFPDVGTYIANPTDT 226
DB 118 NMPGRGAGERLAVDPANSNIIFYFGASGNGLWKSTDGVTFSKVSSFTATGVIIPDPS 177

QY 227 TGYQSDIQGVVWAFKSSSSLSQAQSKTIFVGVADP-NNPVFWSRGGATQWAVPGAPT 285
DB 178 NGVNSDKQGLMWVTFDSTSTTGGATSRIFVGTADNITASVYVSTNAGSTWAVPGQPK 237

QY 286 FIPHKGVFDPVNHVLIATSNTPGVDGSGDVWKFVTSCTWTRISVPSTDTANDYFG 345
DB 238 YFPHKAKLOPAERKALYLTSDGTGPTGLSVWRYDITAGGTWKDITPVSGSL---YFG 294

QY 346 YSLGTLDRQHPNTIMWATQISWPDITIFRGTGATWTRIMDWTSPNRSRLRYLDISA 405
DB 295 FGLGLDLQKPGTLVVASLNSWWPDQAFLPRSTDGTTWSPINAWASYPTETYYISISTPK 354

QY 406 EPLWTFG-VQPNPPVPS---PKLGMWDEAMAI DPFNSRMLYGTGATLYATNDLTWDS 460
DB 355 APWKNFIDVTSESFSDGLIKELGHWIISLEIDPTDSNHWLYGTGWTIFGGHDLTNWDT 414

QY 461 GGQHIAPMKVLEETAANDLISPPSGAPLISALGDLGCFTHA---DVTAVPSTIFTSVP 517
DB 415 RHNVSIOQLADGIEEFSVQDLASAPGSELLAAVGDNDGFTFASRNDLGTSPQWATP 474

QY 518 FTTCTGSDVYAEALNPSIIVRAGSFPDSSQPNDRHVAFTDGGKWFQGSSEPGVVTGGTVA 577
DB 475 WATSTSDVYAGNSVKSVVRVGN-----TAGTQVAISSDGGATWSDIYAADTSMNGGTVA 528

QY 578 ASADGRFVWAPGDPQPVVYAVFGFNSWAASQGVPAANAQIRSDRVNPKTFYALSNGTFY 637
DB 529 YSADGDTILWSTASSG---VQPSQFGSGFASVSSLPAGAVIASDKKTNVFFYAGSGSTFY 585

QY 638 RSTDGGVTFQPVAAAGLPSGAVGM--FHAVPCKEGDLMLAASSGLYHSTNGSSWSAI- 694
Db 586 VSKDTGSSP---TRG-PKLGSAAGTIRIDIAAHPPTAGTLYVSTDVIGIFRSTDSGTTTFQVVS 641
QY 695 TCVSSAVNVGFKSAPGSSYPVAVVGTGGTGA--YESDDCGTTWVLINDDHOYGNW 752
Db 642 TALTNTYQIALGVGS--GSNW-NLYAPGT--GPGSARLYASGSGASWTDIQSGQGFSGID 697
QY 753 QGAITGDHANLRVYGTNGRGIVY--GDIGAPSGSPSPSPSPSPSPSPSPSPSPSSP 810
Db 698 STKVAGSGTAGQVYGTNGRGVGYAQTGGTGT-----SSSTKQSSSSTSSA 748
QY 811 SP 843
Db 749 SSSSTLRSSWSTTRASTVTSSRTSSAAGPTGS 781

RESULT 12
US-10-156-761-10111
; Sequence 10111, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 10111
; LENGTH: 739
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-10111

Query Match 28.1%; Score 1442; DB 14; Length 739;
Best Local Similarity 40.8%; Pred. No. 8.6e-76;
Matches 328; Conservative 107; Mismatches 272; Indels 96; Gaps 25;

QY 7 IRLTWRSRLVSLAATASFAVAALGLVLPITATSPAHAAAT---QPYTWSNVAIGGGG 63
Db 1 MRPSPSRRTV--LACTAA---AALATVP--AVGSOAHAAETTAGPSYRWRNAVIGTG 53
QY 64 FVDGIVFNEGAPGILYVRTDIGNMYRWDANGRWIPLDWWGNVNNYNGVVSIAADPIN 123
Db 54 FVTGVLFPSPVRLAYARTDIGAYRWDGRTPLIDHLGWDNLLGVEAMAVDPHT 113
QY 124 TNKWAACVMTNSDPNDGAILRSSDOGATWQITPLPKLGNMPCRGGERLAVDPNN 183
Db 114 PDRLYLAVGTAAQSWAGN--GAVLRSEDRGATWTRTDLTKLGGNEDGRGAGERLLVDPRD 172
QY 184 DNILYFGAPSGKLMWSTDSGATWSQMTNPPDVGTVIANPTDTTGYQSDIQGVVWAFDK 243
Db 173 STLWLGT--RHDGLLSTDRGATWAAATAPP-----AKANSQGVVF----- 214
QY 244 SSSSLGQASKTIFVGVADPNPNP-----VFWSRDGGATWQAVPGAPTGF----IPHKGVFDP 295
Db 215 ----LVAAGRTVYAGWDGDTSGTANLYRTAD-GTTWGVAPRPSGTSKAVPLRAYDT 269
QY 296 VNHVLIATSNCTGGPVDGSSGDVWKSFSVTGCTWTRISPV-----PSTDTANDYFGYGLTI 351
Db 270 HTRELVTYTGADPGPGQSDGSHKLRATGTATGTWTEVTPVPGGTTSDGSADTFAYGGVAV 329

QY 352 DRQHPNTIMVATQISWMPDTIIIEBSTDGGATWTRIMDWTSYPNBSRLRVYLDISASPMLTF 411
Db 330 DARRPGTLVSTNNRWADGDTVFRSTDGGRTWTSKQ-----AAVEDVSETPFELDW 380
QY 412 GVQPNPVPSPKLGWMDMAIDPFNSDRMLYGTGATLYATNDLTKWDSGGQIHIAPMVK 471
Db 381 GDD-----KPKFGWTLQALAVDPYDSQHVYVGTGATLYGTRDLKRW-----APRIR 426
QY 472 GLEETAVNDLISPPSG--APLISALGDLGGFHADVAVPST-IFTSPTFTTGTSDVYABL 529
Db 427 GLEESAVRQLISPPVGEAHLISGLDIGNVMYHERLTASPSRGMATNPVFGSATGLAQAAA 486
QY 530 NPSIIVPAGSFDSPSQPNDRHVAFSTGGKW--PQSEPGVTTGGTVAASADGSRFW 587
Db 487 RPAYVVRTGMDHGNG-----AYSHDGGRTWAPFEAQPIAKDAPGPIATSDAGTLLW 540
QY 588 A---PGDPGPVVYA---VFGNSWAAASQGVANAIQIRSDRVNPKTFYA--LNGTTFY 637
Db 541 SFVHWDG-----TTYAAHRSTDNGASWSEVSFPKGATPVADPADPTFYAYDFDNGTLY 595
QY 638 RSTDGGVTFQPVAAAGLPSGAVGMFHAVPCKEGDLMLAAS--SGLYHSTNGSSWSAIG 696
Db 596 ASTDSGRSFTARAGLP--SGDSQFKLVAAPCGRSGDLMLSAKWNGLYRSTDDGDTTFARIDS 654
QY 697 VSSAVNVGFKSAPGSSYPVAVVGTGGTTCAYRSDDCGTTWVLINDDHOYGNWQAI 756
Db 655 CWASYTLGFGKAADGADYPAIYQVGTETITAVYRSDDAARTWVRINDDAHQWIGEA 714
QY 757 TGDHANLRVYGTNGRGIVYGD 779
Db 715 VGDPRHGRVLTATNGRGIQYGE 737

RESULT 13
US-10-395-241-12
; Sequence 12, Application US/10395241
; Publication No. US20040038367A1
; GENERAL INFORMATION:
; APPLICANT: YAOI, Katsuro
; APPLICANT: MITSUISHI, Yasushi
; TITLE OF INVENTION: NOVEL XYLOGLYCAN OLIGOSACCHARIDE-DEGRADING ENZYME, POLYNUCLEOTIDE
; TITLE OF INVENTION: ENCODING THE SAME, AND METHOD OF PREPARING THE ENZYME
; FILE REFERENCE: Q73756
; CURRENT APPLICATION NUMBER: US/10/395,241
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: JP 2002-83433
; PRIOR FILING DATE: 2002-03-25
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 12
; LENGTH: 812
; TYPE: PRT
; ORGANISM: Geotrichum sp. M128
US-10-395-241-12

Query Match 21.9%; Score 1125; DB 12; Length 812;
Best Local Similarity 34.5%; Pred. No. 2.8e-57;
Matches 285; Conservative 129; Mismatches 291; Indels 120; Gaps 31;

QY 31 ALGVLPITATSPAHAAATQPYTWSNVAIGGGVFDGIVFNEGAPGILYVRTDIGNMYRW 90
Db 6 SLGKALTALSILASLAVAKHEFPKNAIGGGVITGIVAHPKTKDLYARTDIGGAYRW 65
QY 91 DAANGRWIPLDWWGNVNNYNGVVSIAADPINTKWAAACVMT--NSWDPNDGAILRSS 149
Db 66 DAGTSKWIPLDITEAQDMMNIMGTESALDPNPNDRLYLAQGRYVGDEW----AAYVSE 121
QY 150 DOGATWQITPLPKLGNMPCRGGERLAVDPNNNDNLYFGAPSGKGLWRSTDSGATWSQ 209
Db 122 DRGQSFYIESPPFGANDMGNGRGERLAVNPNFNSVWMTGRT-EGIKWSSDRAKTWTN 180
QY 210 MTNPPDVGTVIANPTDTTGYQSDIQGVVWAFDSSSSSLGQASKTIFVGVADPNPNPVFS 269

Db 181 VTSIPDAF-----TNGIGYTS-----VIFDP-----ERNGTIYASATAPQG-MYVT 220
Qy 270 RDGATWQAVPGAPTGI-----PHKGVPDPVNVHLYIATSNCTGPGY 311
Db 221 HDGGSWEVPAGQPSWLNRTTGAPDKKASIAQPMKVALTP--NFLVTVADYPCPW 278
Qy 312 DGSSGDVWKFESVTSCTWTTRI-----SPVPSDTTANDYFGYGLTITDRQHPNTIMVATQ 364
Db 279 GVTFGEVWRQNRSTGAWDDITPRVGNSSPAPYNNQTFPAGGFCGLSDATNPRLVVIT- 337
Qy 365 ISWMPDTI---IFRSTDDGATWTRIWDTWS-----YPNRSLRVVLDISAEPMLTFG 412
Db 338 LDRDPGALDSIYLSLTDAGATWQVQLSSPNSLGNWGHPTNAARY-KDGTVPVPLDFTN 396
Qy 413 VQP-----NPPVESP---KLGMNDEAMADFPNSDRMLYGTGATLYATNDLTK-----WDS 460
Db 397 NGPQMGYGAPHCTPGTLTKFGWMSAVLIDFPNPEHLMYGTGATWATDLSRVEKDM-- 454
Qy 461 GGOIHIAP---MVKGLETAVNDLISPPSGAPLISALGDLGCTHADVAVPSTIPTSP 516
Db 455 -----APSWYLQIDGIEENAILSLRSPKSGAALLSGIDISGMKHDDLTG-PQKMFAP 507
Qy 517 VFTTGTSDVYAEINPSIIVRAGSFPDSSQPNDRHVAFTDGGKNW--FQSEPGGVTT-- 572
Db 508 QFNLDSIDAAGNFPNVVRAGSSGHEYDSACARGAYATDGGDAWTFPTCPFGMNASHY 567
Qy 573 -GCTVAASADGGRFVWAP--GDGPQPVVYAVFGNSWAASQVPA-----NAQIRSDRV 623
Db 568 QGSTIADVAGSGQIWNSTKLDEQASGPWYSHDYGTWS-----VPAGDLKAQATANVLSDKV 623
Qy 624 NPKTFYALNGTFYRSTDDGVTTPQVAAGLPSSGAVGMFHAVPGKBDLWL-AASSGLY 682
Db 624 QDCTFYATDGGKFFVSTDDGKSYAAKAGGLVT--GTSMLPAPVNPVWAGDVWVPVEGGLF 681
Qy 683 HSTNGSSWSAI-TGVSSAVNVGFGKS-----APGSSYPVAVFVGT--IGVTVGAYRSDDC 735
Db 682 HSTDFGASFRVGTANATLVSVGAPKSKDGKASAPSAVFIWGTDKPGSDIGLYRSDDN 741
Qy 736 GTTWLVINDHOYGNWQQAITGDHANLRVYIGTNGRGIVYGD 780
Db 742 GSTWTRVNDQEHNSYSG-PTWIEADPKVYGRVYLTNGRGIVYADL 785

RESULT 14
US-10-395-241-14
; Sequence 14, Application US/10395241
; Publication No. US20040038367A1
; GENERAL INFORMATION:
; APPLICANT: YAOI, Katsuro
; APPLICANT: MITSUISHI, Yasushi
; TITLE OF INVENTION: NOVEL XYLOGLUCAN OLIGOSACCHARIDE-DEGRADING ENZYME, POLYNUCLEOTIDE
; FILE REFERENCE: Q73756
; CURRENT APPLICATION NUMBER: US/10/395,241
; PRIOR FILING DATE: 2003-03-25
; PRIOR FILING DATE: 2002-03-25
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 14
; LENGTH: 789
; TYPE: PRT
; ORGANISM: Geotrichum sp. M128
US-10-395-241-14

Query Match 21.8%; Score 1117; DB 12; Length 789;
Best Local Similarity 35.0%; Pred. No. 7.9e-57;
Matches 281; Conservative 124; Mismatches 279; Indels 120; Gaps 31;
Qy 52 YWNSNVAICGGFVGDIVFNEGAPGLIYVRTDIGMYRWDANANGRWIPLLDWYGNWGY 111
Db 4 YEPKNAIGGGVITGIVAHPKTKDLLYARTDIGGAYRWDAGTSKWIPLNDFEADQDMNI 63

Qy 112 NGVVSIAADPINTNKVAAVGYMT-NSWDPNDGAILRSSDQATWQITPLPKLGNMPC 170
Db 64 MGTESIALDPNPDRLYLAAQGRYVGDEW-----AAFVSEDRCQSFYIYSSPFPMGANDMG 119
Qy 171 RGMERLAVDPNNNLLYFGAPSGKGLWRSTDSGATWSOMTNFDPDVGTVIANPTDTTGYQ 230
Db 120 RANGERLAVNPNSEWVMGTRT-EGIKSSDRAKTWTNVTISIPDAF-----TNGIGYT 172
Qy 231 SDIQGVVWVAFPKSSSLSGQASKTIFVGVADPNPNFVMSRDGGATWQAVPGAPTGI--- 287
Db 173 S-----VIFDP-----ERNGTIYASATAPQG-MYVTHDGGVSWEPVAGQPSWLNRT 218
Qy 288 -----PHKGVPDPVNVHLYIATSNCTGPGYGLTITDRQHPNTIMVATQISWMPDTI---IFRSTDDGAT 382
Db 219 TGAPDKKASIAQPMKVALTP--NFLVTVADYFGPVGWTFGEVWRQNRSTGAWDDIT 276
Qy 332 -----SPVPSDTTANDYFGYGLTITDRQHPNTIMVATQISWMPDTI---IFRSTDDGAT 382
Db 277 PRVGNSSPAPYNNQTFPAGGFCGLSDATNPRLVVIT-LDRDPGALDSIYLSLTDAGAT 335
Qy 383 WTRIWDTWS-----YPNRSLRVVLDISAEPMLTFGVQP-----NPPVESP---KL 425
Db 336 WKDVTQLSSPNSLGNWGHPTNAARY-KDGTVPVPLDFTNNGPQMGYGAPHCTPGTLTKFG 394
Qy 426 WMDAMADFPNSDRMLYGTGATLYATNDLTK---WDSGGQIHIAP---MVKGLEETA 477
Db 395 WMSAVLIDFPNPEHLMYGTGATWATDLSRVEKDM-----APSWYLQIDGIEENA 446
Qy 478 VNDLISPPSGAPLISALGDLGCTHADVAVPSTIPTSPVFTTGTSDVYAEINPSIIVRA 537
Db 447 ILSLRSPKSGAALLSGIDISGMKHDDLTG-PQKMFAPQFNLDSIDAAGNFPNVVRA 505
Qy 538 GSPDPSQPNDRHVAFTDGGKNW--FQSEPGGVTT--GCTVAASADGGRFVWAP--G 590
Db 506 GSSGHEYDSACARGAYATDGGDAWTFPTCPFGMNASHYQGSTIADVAGSGQIWNSTKLD 565
Qy 591 DPGQPVVYAVFGNSWAASQVPA-----NAQIRSDRVNPKTFYALNGTFYRSTDDGV 644
Db 566 EQASGPWYSHDYGTWS-----VPAGDLKAQATANVLSDKVQDCTFYATDGGKFFVSTDDGK 621
Qy 645 TFQPVAAGLPSSGAVGMFHAVPGKBDLWL-AASSGLYHSTNGSSWSAI-TGVSSAVN 702
Db 622 SYAAKGAGLVT--GTSMLPAPVNPVWAGDVWVPVEGGLFHSSTDFGASFRVGTANATLVS 679
Qy 703 VGFGKS-----APGSSYPVAVFVGT--IGVTVGAYRSDDC GTTWLVINDHOYGNWQQA 756
Db 680 VGAPKSKDGKASAPSAVFIWGTDKPGSDIGLYRSDDN GSTWTRVNDQEHNSYSG-PTWI 738
Qy 757 TGDHANLRVYIGTNGRGIVYGD 780
Db 739 EADPKVYGRVYLTNGRGIVYADL 762

RESULT 15
US-10-395-241-18
; Sequence 18, Application US/10395241
; Publication No. US20040038367A1
; GENERAL INFORMATION:
; APPLICANT: YAOI, Katsuro
; APPLICANT: MITSUISHI, Yasushi
; TITLE OF INVENTION: NOVEL XYLOGLUCAN OLIGOSACCHARIDE-DEGRADING ENZYME, POLYNUCLEOTIDE
; FILE REFERENCE: Q73756
; CURRENT APPLICATION NUMBER: US/10/395,241
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: JP 2002-83433
; PRIOR FILING DATE: 2002-03-25
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 18
; LENGTH: 826
; TYPE: PRT
; ORGANISM: Artificial Sequence

FEATURE:
; OTHER INFORMATION: Xyloglucan Oligosaccharide-Degrading Enzyme with Histidine Tag
US-10-395-241-18

Query Match 21.7%; Score 1114; DB 12; Length 826;
Best Local Similarity 35.0%; Pred. No. 1.2e-56;
Matches 281; Conservative 123; Mismatches 280; Indels 120; Gaps 31;

QY 52 YTSNVAIGGGFVDGIVNEGAGILYVRTDIGMYRMDAANGRWIPLLDWVGNNWGY 111
D 5 YEFKNAVIGGGYITGIVAHPKTKDLLYARTDIGGAYRWDAGTSKWIPLNDFIEAQDMNI 64

QY 112 NGVVSIAADPINTKVAAGMT-NSWDPNDCAILRSSDQATWITPLPFKLGNNMPG 170
D 65 MGTESIALDPNNDRLYLAQGRVGDW-----AAFTVSEDRGQSFTIYESPPFMGANDMG 120

QY 171 RGMGERLAVDPNNNTILYFCGAPSGKGLWRSTDSGATWSQMTNPPDVGYIANPTDITGYQ 230
D 121 RANGERLAVNPFNSNEVMGTRT-EGIKSSDRAKTWNTVTSIPDAP-----TNGIGYT 173

QY 231 SDIQGVVWVAFDKSSSSSLGQASKTIFVGADPNPNPVFWSRDGGATWQAVEGAPTGI--- 287
D 174 S-----VIFDP-----ERNGTIYASATAPOG-MYVTHDGGVSWEPVAGQPSSMLNRT 219

QY 288 -----PHKGVDPNVHVIYIATNTGGPYDGGSSGDVWKFSTSGTWTRI- 331
D 220 TGAFFDKKPAIAPOQPMKVALTE--NFLYVYADYPCPMGVTFKVMRQNRRTSGAWDDIT 277

QY 332 -----SPVPSTANDYFCYSGLTIDRQHPNTIMVATQISWHPDTI---IPFSTDGGAT 382
D 278 PRVGNSSPAPYNNQTPPAGFCGLSDVATNPRLVWIT-LDRDPGALDSIYLSTDAGAT 336

QY 383 WTRINDWTS-----YENRSLRYVLDISAEPWLTTFGVQP-----NPPVPSP---KLG 425
D 337 WKDVTQLSSPNSLEGNWCHPTNAARY-KDGTVPVWLDFNNGPQWGGVGAPHGTFGLTKFG 395

QY 426 WNDKMAIDPFNSDRMLYGTGATLYATNDLTK-----WDSGQIHIAP-----MYKGLEETA 477
D 396 WMSAVLIDPFNEHLMYGTGATWATDLSRVEKDW-----APSWYLQIDGIEENA 447

QY 478 VNDLISPPGAPLISALDGLGGTHADVTAVPTFTSPTFTTGTSDVYAEINPSIIVRA 537
D 448 ILSLRSPKSGAALLSGIGDISGMKHDLLTK-PQKMFCAPOFSLNDSIDAAGNFPNVVVRA 506

QY 538 GSPDPSSQPNDRHVAFSTDCGKW--FQSGEPGGVTT---GGTVAASADGSRFVWAP--G 590
D 507 GSGHEYDSACARGATATDGDATWITPTCPGWNASHYQGSTIAVDASGSIWSTKLD 566

QY 591 DPQQPVVYAVFGNSWAASQGVPA-----NAQIRSDRVNPKTFYALNSGTFYRSTDDGV 644
D 567 EQASGPWYSHDYKTSW----VPAGDLKAQTANVLSDKVODGTFYATDGGKFFVSTDDGK 622

QY 645 TPQPVAAGLPSSGAVGVWFHAFVCKEGDLWL-AASSGLYHSTNGGSSWSAI-TGVSSAVN 702
D 623 SYAAKAGGLVT--GTSLMPAVNFWAGDVWVPVPEGGLFHSTDFGASFTRVGTANATLVS 680

QY 703 VGFQKS-----APGSSYPAVFVGT--IGGYTGAYRSDCGTWTWLINDDOHQYGNMGQAI 756
D 681 VQAPKSKDGKASAPSAFVGTGDKGSDIGLYRSDDNGSTWTRVNDQEHNYSG-PTMI 739

QY 757 TGDHANLRRVYIGTNGRGIVYGDI 780
D 740 EADPKVYGVYLTNGRGIVYADL 763

Search completed: May 14, 2004, 09:24:15
Job time : 58 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 14, 2004, 08:53:04 ; Search time 9066 Seconds
(without alignments)
4575.257 Million cell updates/sec

Title: US-09-917-376-1
Perfect score: 5135
Sequence: 1 MDRSENIRLWMSRRLVSL.....RASFGSVNPATPTADTYLOX 957

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 3470272 seqs, 2167151695 residues
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US09917376/runat_14052004_085256_4619/app_query.fasta.1.1095
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=oct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pf0 -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09917376.@CGN_1_1_5932@runat_14052004_085256_4619 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -XGAPOP=10 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:*

- 1: gb_ba:*
- 2: gb_htg:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sta:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vl:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_om:*
- 21: em_ox:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_ph:*
- 25: em_pl:*
- 26: em_ro:*
- 27: em_sta:*
- 28: em_un:*

- 29: em_vi:*
- 30: em_htg_hum:*
- 31: em_htg_inv:*
- 32: em_htg_other:*
- 33: em_htg_mus:*
- 34: em_htg_pln:*
- 35: em_htg_rod:*
- 36: em_htg_mam:*
- 37: em_htg_vrt:*
- 38: em_sy:*
- 39: em_htgo_hum:*
- 40: em_htgo_mus:*
- 41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	5134	100.0	2869	6	AX700058	AX700058 Sequence
2	2493	48.5	296500	1	SC039128	AL939128 Streptomy
3	2487	48.4	299800	1	AP005028	AP005028 Streptomy
4	2419.5	47.1	3262	1	AF078038S1	AF078038 Caldicell
5	2174	42.3	2823	6	AX565635	AX565635 Sequence
6	2037	39.7	12732	1	AE007608	AE007608 Clostridi
7	2007.5	39.1	2950	1	AJ585344	AJ585344 Clostridi
8	1784	34.7	2579	8	AB1292929	AJ292929 Agaricus
9	1782.5	34.7	2745	8	AB125511	AB125511 Aspergill
10	1690	32.9	3959	8	AY040839	AY040839 Aspergill
11	1633	31.8	2724	8	AK110506	AK110506 Oryza sat
12	1631.5	31.8	2517	8	AY281371	AY281371 Hypocrea
13	1570	30.6	299550	1	AP005031	AP005031 Streptomy
14	1179.5	23.0	14520	1	AE011809	AE011809 Xanthomon
15	1170.5	22.8	11910	1	AE012276	AE012276 Xanthomon
16	1161	22.6	13278	1	AE001712	AE001712 Thermotog
17	1125.5	21.9	2646	6	AX924429	AX924429 Sequence
18	1125.5	21.9	2716	8	AB089343	AB089343 Geotrichu
19	1124	21.9	2481	6	AX924435	AX924435 Sequence
20	1117	21.8	2367	6	AX924431	AX924431 Sequence
21	823	16.0	3687	6	AX700036	AX700036 Sequence
22	702	13.7	2289	6	AX700050	AX700050 Sequence
23	460.5	9.0	590	8	AB1534359	AB1534359 Agaricus
24	431.5	8.4	5437	1	CASR69XYN2	AF036924 Caldicell
25	429	8.4	3365	6	AX700025	AX700025 Sequence
26	409	8.0	11707	1	AF078737	AF078737 Caldicell
27	409	8.0	11707	6	E35099	E35099 Truncated c
28	407	7.9	299175	1	AP005023	AP005023 Streptomy
29	402.5	7.8	3237	1	AF200304	AF200304 Caldibaci
30	398.5	7.8	5513	1	AT286105	Z86105 A.thermophi
31	397.5	7.7	5439	1	CDCCELA	L32742 Caldicoellum
32	391.5	7.6	5284	1	CSU16308	U16308 Caldicoellum
33	389	7.6	24208	1	AY28135782	AY281358 Unculture
34	387	7.5	344615	1	EX569695	EX569695 Synchoco
35	381.5	7.4	2801	5	AF218784	AF218784 Gallus ga
36	380	7.4	209091	10	AC122299	AC122299 Mus muscu
37	372.5	7.3	2015	8	CREVSP3A	L29029 Chlamydomon
38	370	7.2	206849	2	AC011760	AC011760 Drosophil
39	367.5	7.2	215910	10	AC127335	AC127335 Mus muscu
40	367	7.1	7085	2	AC015260	AC015260 Drosophil
41	367	7.1	179386	3	AC104515	AC104515 Drosophil
42	367	7.1	273948	3	AE003574	AE003574 Drosophil
43	365	7.1	3768	2	AC015261	AC015261 Drosophil
44	365	7.1	117167	9	AC110813	AC110813 Homo sapi
45	365	7.1	240904	2	AC126204	AC126204 Rattus no

ALIGNMENTS

RESULT 1

AX700058
 LOCUS AX700058 2869 bp DNA linear PAT 03-APR-2003
 DEFINITION Sequence 2 from Patent WO03012090.
 ACCESSION AX700058
 VERSION AX700058.1 GI:29536021
 KEYWORDS
 SOURCE Acidotherrnus cellulolyticus
 ORGANISM Acidotherrnus cellulolyticus
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 Frankineae; Acidotherrmaceae; Acidotherrmus.
 REFERENCE
 1 Ding, S.Y., Adney, W.S., Vinzant, T.B. and Himmel, M.E.
 Thermal tolerant avicelase from *Acidotherrmus cellulolyticus*
 Patent: WO 03012090-A 2 13-FEB-2003;
 JOURNAL Midwest Research Institute (US)
 FEATURES
 Location/Qualifiers
 1..2869
 /organism="Acidotherrmus cellulolyticus"
 /mol_type="unassigned DNA"
 /db_xref="taxon:28049"
 ORIGIN
 Alignment Scores:
 Pred. No.: 3,44e-147 Length: 2869
 Score: 5134.00 Matches: 956
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 99.98% Indels: 0
 DB: 6 Gaps: 0
 US-09-917-376-1 (1-957) x AX700058 (1-2869)
 QY 1 MetAspArgSerGluAsnIleArgLeuThrMetArgSerArgValSerLeuLeu 20
 DB 1 ATGGATCGTTCGGAGAACATCCGCTCTGACTATGAGATCAGACATTTGGTATCACTGCTC 60
 QY 21 AlaAlaThrAlaSerPheAlaValAlaAlaLeuGlyValLeuProIleAlaIleThr 40
 DB 61 GCGGCCACTCGGTCGTTGCGCGTGGCGCGCTCTGGAGTCTCTGCCATCGCGATAAG 120
 QY 41 AlaSerProAlaHisAlaAlaThrThrGlnProTyrThrTrpSerAsnValAlaIleGly 60
 DB 121 GCTTCTCTCGGCACGCGCGAGACTCAGCCGTACACCTGGAGCAACGTTGGCGATCGGG 180
 QY 61 GlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrVal 80
 DB 181 GCGCGCGCTTTTGTCCAGCGGATCGTCTTCAATGAAGTGCACCGGAATCTGTACGTG 240
 QY 81 ArgThrAspIleGlyGlyMetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeu 100
 DB 241 CGGACGGACATCGGGGGATGTATCGATGGATGCCCAACGGCGGTGGATCCCTCTT 300
 QY 101 LeuAspTrpValGlyTrpAsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAsp 120
 DB 301 CTGGATTGGGTGGATGGAAATTTGGGGTCAACCGCGCTCGTCAAGCTTGGCGGACAG 360
 QY 121 ProIleAsnThrAsnIleValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspPro 140
 DB 361 CCGATCAATACTAACAGGTATGGCGCGCTCGGAATGTACCAACAGCTGGGACCCA 420
 QY 141 AsnAspGlyValAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeu 160
 DB 421 AACGACGGAGCGATTTCTCCGCTCTGTATCAGGGCGCAACGTGGCAATAACGCCCTTG 480
 QY 161 ProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAsp 180
 DB 481 CCGTTCAGGCTTGGCGCAACATGCCCGCGTGGAAATGGCGAGCGGCTTGGGTGGAT 540
 QY 181 ProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSer 200
 DB 541 CCAAAACATGACAAACATTTCTGTTATTTTCGGCGCGCCCGAGCGCGCAAGGGCTCTGGAGAAGC 600
 QY 201 ThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyrIle 220

DB 601 ACAGATTCCCGCGGACCTGTGTCAGATGACGAACCTTTCGGACGTAGGACGACGTACATT 660
 QY 221 AlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAla 240
 DB 661 GCAATCCCACTGACACGACCGGCTATCAGAGCGATATTCAGGCGTCTGTGGGTGGT 720
 QY 241 PheAspLysSerSerSerSerLeuGlyGlnAlaSerLysThrIlePheValGlyValAla 260
 DB 721 TTCGACAAGTCTTCTGTCATCGTCGGGCAAGCGAGTAAGACCATTTTGTGGCGTGGCG 780
 QY 261 AspProAsnAsnProValPheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValPro 280
 DB 781 GATCCCAATAATCCGCTTCTTGAGCAGACAGACGCGCGCGAGCTGGCAGCGGTGGCG 840
 QY 281 GlyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsnHisValLeu 300
 DB 841 GGTGCGCGGACCGGCTTCATCCCGACAGAGGCGCTCTTTGACCGGTCAACACGCTGCTC 900
 QY 301 TyrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLys 320
 DB 901 TATATTGCCACACGCAATACGGGTGGTCCGATGACGGGAGCTCCGGCGACGCTCGGAAA 960
 QY 321 PheSerValThrSerGlyThrTrpThrArgIleSerProValProSerThrAspThrAla 340
 DB 961 TTCTCGGTGACCTCCGGGACATGACGCGAATCAGCCGCTACCTTCGACGCGACACGGCC 1020
 QY 341 AsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMet 360
 DB 1021 AACGACTACTTTGGTTACAGCGGCTCTACTATGACCGCCAGCACCCCAACACGATAATG 1080
 QY 361 ValAlaThrGlnIleSerTrpTrpProAspThrIleIlePheArgSerThrAspGlyGly 380
 DB 1081 GTGGCAACCCAGATATCGTGGTGGCGGACACCATATCTTTTCGGAGCACCGCGCGT 1140
 QY 381 AlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrVal 400
 DB 1141 GCGACGTGGACGCGGATCTGGGATTTGGAGAGTTATCCCAATCGAAGCTTGGATATGTG 1200
 QY 401 LeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnProProValPro 420
 DB 1201 CTTGACATTTTCGGCGAGCTTGGTGTACCTTCGGCGTACAGCCGAATCTCCCCGTACC 1260
 QY 421 SerProLysLeuGlyTyrMetAspGluAlaAlaMetAlaIleAspProPheAsnSerAspArg 440
 DB 1261 AGTCCGAAGCTCGGTGGATGGATGAGCGATGGCAATCGATCCGTTCAACTCTGATCG 1320
 QY 441 MetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSer 460
 DB 1321 ATGCTCTACGGAACAGCGCGACGTTGTACGCAACAAATGATCTCAGAAATGGGACTCC 1380
 QY 461 GlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluThrAlaValAsnAsp 480
 DB 1381 GCGCGCGAGATTCATATTCGCGCCGATGGTCAAGGATTTGGAGGAGACGCGCGTAACGAT 1440
 QY 481 LeuIleSerProProSerGlyValProLeuIleSerAlaLeuGlyAspLeuGlyGlyPhe 500
 DB 1441 CTATCAGCCCGCGCTTGGCGCCCGCTCATCAGCGCTCTCGAGAGACTCTCGCGGCTTC 1500
 QY 501 ThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProValPheThrThr 520
 DB 1501 ACCCACGCGGAGTACTTCCGCTGCCATCGACGATCTTTCACGTCACCGGTGTTACGACC 1560
 QY 521 GlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPhe 540
 DB 1561 GGCAACGCGTGCATATTCGCGAATTTGAATCCGTCGATCATCGTTCGCGCTCGGAAGTTTC 1620
 QY 541 AspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsn 560
 DB 1621 GATCCATCGAGCCCAACCGGACGAGCGGACGCTCGGCTTCTCGACAGACGCGCGCAAGAC 1680
 QY 561 TrpPheGlnGlySerGluProGlyValThrThrGlyGlyThrValAlaAlaSerAla 580

Db 1681 TGGTTCAGGCGAGCAACCTCGCGGGGTGACGAGCGGGCGACCCGTGCGCGCATCGGCC 1740

Qy 581 AspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValTyrAlaVal 600
|||||

Db 1741 GACGGCTCTGTTCTGCTGGGCTCCCGCGCATCCCGTCAAGCTGTGGTGTAGCGAGTC 1800

Qy 601 GlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAlaGlnIleArgSer 620
|||||

Db 1801 GGATTTGGCAACTCTCTGGGCTGCTTCGCAAGGTGTTCGCCCAATGCCAGATCCGCTCA 1860

Qy 621 AspArgValAsnProLysThrPheTyrAlaLeuSerHsnGlyThrPheTyrArgSerThr 640
|||||

Db 1861 GACCGGTGAATCCAAAGACTTTCTATGCTCTATCCCAATGAACTTCTATCAAGGACG 1920

Qy 641 AspGlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGly 660
|||||

Db 1921 GACGGGGGTGNACATCCACCGTTCGCGCGGCTCTTCGAGCAGCGTGGCTCGT 1980

Qy 661 ValMetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaSerSerGly 680
|||||

Db 1981 GTCATGTTCCAGCGGTCTCGAAAGAGCGATCTGTGGTCTGCTGCATCGAGCGG 2040

Qy 681 LeuTyrHisSerThrHsnGlySerSerTrpSerAlaIleThrGlyValSerSerAla 700
|||||

Db 2041 CTTTACCACTCAACCAATGCGCGCAGCTGTGCTGCAATCAACCGCGTATCTCCGCG 2100

Qy 701 ValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheValVal 720
|||||

Db 2101 GTGAACGTGGGATTTGGTAAGTCTGCGCGCGGTGCTCATACCGAGCGGTCTTGTGCTG 2160

Qy 721 GlyThrIleGlyValThrGlyAlaTyrArgSerAspAspCysGlyThrTrpVal 740
|||||

Db 2161 GGCACGATCGAGCGCTTACGGGGGCTACCGTCCGACGACTGTGGGACGACTGGGTA 2220

Qy 741 LeuIleAsnAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHis 760
|||||

Db 2221 CTGATCAATGATGACGACCAATACGGAATTTGGGACAGCAATCAACCGGTGACCA 2280

Qy 761 AlaAsnLeuArgArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIle 780
|||||

Db 2281 GCGAATTTACGGCGGTGTACATAGGACAGCAACGCGCGTGGAAATGTATACGGGGACAT 2340

Qy 781 GlyGlyAlaProSerGlySerProSerProSerValSerProSerAlaSerProSerLeu 800
|||||

Db 2341 GGTGGTGGCGCTCCGATCGCGCTCTCCGTCGGTGAATCCGTGCGCTTCGCGAGCGCTG 2400

Qy 801 SerProSerProSerProSerProSerProSerProSerProSerProSerProSerPro 820
|||||

Db 2401 AGCCGAGCCGAGCCGAGCAGCTCGCCATCGCGTCCGCGTCCGCGAGCTCGAGTCCA 2460

Qy 821 SerSerSerProSerProSerProSerProSerProSerProSerProSerProSerPro 840
|||||

Db 2461 TCTCTGTCGCGCTCTCCGTCGCGCTCACCATCCGCGAGTCCGCTCGCTCCGTCACCA 2520

Qy 841 SerAlaSerProSerProSerProSerProSerProSerProSerProSerProSerPro 860
|||||

Db 2521 TCGCGCTCCCGAGCCGCTCTTCGTCAGGCGCGCTTCGTCACCGCTTCGTCGCGG 2580

Qy 861 SerProThrProSerSerProValSerGlyGlyValLysValGlnTyrLysAsnAsn 880
|||||

Db 2581 AGCCCAACCGCTGCTGCTCGCGGTGTCCGGTGGGGTGAAGTGCAGTATAAGATAAT 2640

Qy 881 AspSerAlaProGlyAspAsnGlnIleLysProGlyLeuGlnValValAsnThrGlySer 900
|||||

Db 2641 GATTCGGCGCGGGTGATTAATCAGATCAAGCCGGGTTTGCAGGTGTGAATACCGGGTCG 2700

Qy 901 SerSerValAspLeuSerThrValThrValTyrTrpPheThrArgAspGlyGlySer 920
|||||

Db 2701 TCGTCGGTGGAATTTGTCGACGGTGACGGTCCGGTACTGTGGTTTCAACCGGGATGTGGCTG 2760

Qy 921 SerThrLeuValTyrAsnCysAspTrpAlaAlaIleGlyCysGlyAsnIleArgAlaSer 940
|||||

Db 2761 TCACACTGTTGACAACTGTGACTGGCGCGCATCGGGTGTGGGAATATCCGCGCTCG 2820

Qy 941 PheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 956
|||||

Db 2821 TTCGCTCGGTGAACCCGCGACGCGCGGCGGACACCTACCTCGCAG 2868

RESULT 2

SC0939128/c 296500 bp DNA linear BCT 11-FEB-2003

LOCUS Streptomyces coelicolor A3(2) complete genome; segment 25/29.

DEFINITION AL039128 AL032862 AL031013 AL031107 AL031350 AL031371 AL031515

ACCESSION AL033505 AL034443 AL035161 AL079355 AL096825 AL645882

VERSION AL939128.1 GI:24413907

KEYWORDS Streptomyces coelicolor A3(2)

SOURCE Streptomyces coelicolor A3(2)

ORGANISM Bacteria: Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

REFERENCE 1 Bentley, S.D., Chater, K.F., Cerdeno-Tarraga, A.M., Challis, G.L., Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H., Harper, D., Batenan, A., Brown, S., Chandra, G., Chen, C.W., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S., Huang, C.H., Kieser, T., Larke, L., Murphree, L., Murphy, K., O'Neill, S., Rabinowitsch, E., Rajandream, M.A., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Wietzorrek, A., Woodward, J., Barrall, B.G., Parkhill, J. and Hopwood, D.A.

TITLE Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2)

JOURNAL Nature 417 (6885), 141-147 (2002)

MEDLINE 21996410

PUBMED 12000953

REFERENCE 2 (bases 1 to 296500)

AUTHORS Bentley, S.D.

TITLE Direct Submission

JOURNAL Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces sequencing team, Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: sdb@sanger.ac.uk

COMMENT On or before Oct 26 2002 this sequence version replaced GI:20520750, GI:20520810, GI:20520753, GI:20520666, GI:20520757, GI:20520758, GI:20520668, GI:20520759, GI:20520814, GI:20520765, GI:20520767.

FEATURES

source 1. 296500

/organism="Streptomyces coelicolor A3(2)"

/mol_type="genomic DNA"

/strain="A3(2)"

/db_xref="taxon:100226"

complement (144..443)

/gene="SCO6480"

/note="synonym: SC9C7.16c"

complement (144..443)

/gene="SCO6480"

/note="SC9C7.16c, unknown, len: 99 aa"

/codon_start=1

/translation="hypothetical protein SC9C7.16c"

/protein_id="CAA22728.1"

/db_xref="GI:4154075"

/db_xref="SPTREMBL:Q92B04"

/translation="MTYGCIAASMKTGPGFRERVSLLLSAADGLRDAGCELYVVGLE DDETTIWTVEVTRDHDASLELPAAKDAIGRAMPMLTGETFKQEMSVAGGLG"

complement (517..1563)

/gene="SCO6481"

/note="synonym: SC9C7.17c"

complement (517..1563)

/gene="SCO6481"

/note="SC9C7.17c, conserved hypothetical protein, len: 348aa; similar to several hypothetical proteins eg. SW:Y0BN MYCTU from Mycobacterium tuberculosis (340 aa) fasta scores; opt: 998, z-score: 1321.8, E(): 0, (47.6% identity in 355 aa overlap)."

/codon_start=1

```

/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAA22729.1"
/db_xref="GI:4154076"
/db_xref="SPTREMBL:Q9ZBJ3"
/translation="WRWENLTSDHARAADTALFGADAVVTEFTDPEFGITFEHVR
ARSLNRVPGASRPPEWTVNFRGTHACVYCFARKTHSYLDLDTGLGFGDSQIVVVK
NAPELVRLQASRWGQEVAMGTNVCYORAGRYRLMFGPIIEALTERANPFSILTK
GTLILDLRLTRAAVTVSGISVSGFTDABLWRTVFPPTAPERLLEFVWALGGRG
ICGVLMAPIVIFLGDPEAQLRAIVAAAGTSVTLVHLRPGAREFWMAWLGRRH
HPLVRYERYADGAYAKWYQRRITRQVHDLAREYIGIGPARAGMRPRVAPEHSRP
AEPAPPEPQLSLI"
1800..2243
/gene="SCO6482"
/notes="SC9C7.18"
/notes="synonym: SC9C7.18"
1800..2243
/gene="SCO6482"
/notes="SC9C7.18, conserved hypothetical protein, len: 348
aa; similar to two hypothetical proteins e.g. TR:069685
(EMBL:AL022121) Mycobacterium tuberculosis RV03718c or
MTV025.066c, 147 aa; fasta scores; opt: 475, z-score:
768.3, E(): 0, (52.1% identity in 146 aa overlap)"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAA22730.1"
/db_xref="GI:4154077"
/db_xref="SPTREMBL:Q9ZBJ2"
/translation="MAQVETATRIVAADAETVFDLADYSGTRELKLEHSEYEV
EGDGGEGTLVHWKLAISVRKRDCLLEVTEDTDELVDKDRNSMTWTRVTPAGEGK
SRVVLVTWQAGGVGGFFKTFAPKGLRIYDALLAKLAEEVK"
2394..4841
/gene="SCO6483"
/notes="SC9C7.19"
/notes="synonym: SC9C7.19"
2394..4841
/gene="SCO6483"
/notes="SC9C7.19, probable efflux protein, len: 815aa;
similar to many, several of which are involved in toxin
/antibiotic efflux eg. TR:069070 (EMBL:AF061335)
oxytetracycline exporter from Streptomyces rimosus (536
aa) fasta scores; opt: 1020, z-score: 879.9, E(): 0,
(36.9% identity in 542 aa overlap)"
/codon_start=1
/transl_table=11
/product="putative efflux protein"
/protein_id="CAA22731.1"
/db_xref="GI:4154078"
/db_xref="SPTREMBL:Q9ZBJ1"
/translation="MGGITLVQDPVDPVQDEPVVAQAQPPGPPVPDPGERDAGLSP
RRVRLVFGMLALLALEOMIVATLPIKIVGELHGLDKMSWAIYVLLTSTVGLPI
YKGLDLGKGVQFQAILVFFVGSALAGRAQTMOLIAFRAVQGVGAGLMTGVQAI
IADIVPRERARKMGLIGASFGLASVAGPELLGTYFDHLNWRWCFFYINPFGLVTA
VAVLPLPRVRKPRDILGALLAAASTCLVLLTWSGGTEYAWGSRVILGLAAGVA
ATLFLAEPHFAPELPIRLFRDSIFNTALVGLVGVVALFCAASLYLFLQMVDGA
SATESGLLPMWGVGASIIISGLISRTHYRHPILGSLSVWMLLSRLDADT
SRLYSIMWALVLAGLGVMPVLAVONSVRPTDGTATSNANYPROIGSGVGAFF
GTLFAGRLDALADIPAEAGVGLPAEALITQLVHSLPALRDYAGVADAMPRI
LYLVPLVGLGLFALKKPKVNSHNAVPTDDEPAFQPPQPAQPPAPSPFAGGI
PVCVTQHPDGTVPVPAALTLIDVGGQIGRGASGDDGYALATPGSGAYVLTAAAG
HOQAVSVTVGRPVLDVLVLAGRAGLSVMTADGSPVRDATVTLTNVHGVEATTR
SGREGYVITELVAGYVTLAAGAPAPPAALPVSOAARETRODVLAGGAVLRGTVR
ASGGRVEDARVTLDAAGNVDTLTTPGDFGRFVLDLSSGEYTVIAAGYPPVATVLQ
VAGGGERDLQLGHED"
4911..7493
/gene="SCO6484"
/notes="SC9C7.20"
/notes="synonym: SC9C7.20"
4911..7493
/gene="SCO6484"
/notes="SC9C7.20, conserved hypothetical protein, len:
860aa; similar to many hypothetical proteins egs.
TR:086875 (EMBL:U22894) from Streptomyces lividans (547
aa) fasta scores; opt: 1000, z-score: 738.2, E(): 0,

```

```

(40.0% identity in 497 aa overlap) and TR:086807
(EMBL:AL031031) from Streptomyces coelicolor (916 aa)
fasta scores; opt: 358, z-score: 316.9, E(): 2.3e-10,
(27.2% identity in 779 aa overlap). Also contains a short
region of similarity to TR:Q53897 (EMBL:X60316) Aaa
regulatory locus from Streptomyces coelicolor (192 aa)
fasta scores; opt: 108, z-score: 223.2, E(): 3.8e-05,
(38.5% identity in 135 aa overlap)".
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAA22732.1"
/db_xref="GI:4154079"
/db_xref="GOA:Q9ZBJ0"
/db_xref="SPTREMBL:Q9ZBJ0"
/translation="MVTGHPRTVVVTAQILRSRGQRGPDAMDHGTEDAPPGHAGP
TVPADPVTGRVPLAVVDRDGLVSHWSRGARRLFSKEAIGAPAPDLLPVSGVLP
EDAEPCAGHDDSYRSYTAHDGPGPGLESSLDGRLSYPAGRABLTVRADRDVILW
ARPLVPGPERLLVLAADAGLRDYADGDDSGAPVVERIAFAFALHTDFFPGDEL
AYPLPILPMSVGSARIQVLELGPVLEFSQHDVPVTPDWGPRRAERARRE
RAAALAAAGEPVPDELDEADEADLEAARERLEFNEVSGAIGTSLDLSRTIVEVRA
VVPRTDVAGTYLREOVVAGEGPDGVPDVTMMHVRVALEHTDPRGWDVVPVGEAM
PFAHTPFQCTTGTGEPLVPRITEEMGHAIASOFEKDIPRLITGRSMLVPLKART
VULGFMLLRHPRPVPNDRVTCASLARAGLVLDNARMYTHQENVAEITLQDSMLP
HIPARLAGCDIATRYLPGLTLLRGVGGWFDKLPKARTALVVGDMGHGNSAAMMG
QLRTAVQTWAGLDLPPAQLRLNDLQALRGDTHLATCLXAVYDPIAGELHLNAGHI
PVLVRAEDGASBLIDLPTGAPIGVGGVPFESVRVTVPAGDRLVCTDGLVEVRGDI
GVLATLCEASAAHPAAMDACDTIIRALGAGGKDDVALLMARLTGIEPDVAEMR
DLDAEPAVGEARAAREOLYDMGLPLVAPALGMVSELVTNVAHSHARPVALRLVRA
DTLLCEVDDDHLELPALKSAGPDGTGRGLRVVSTLAREWGSRTGAGKTWNFELTLP
ARRH"
7711..8433
/gene="SCO6485"
/notes="SC9C7.21"
/notes="synonym: SC9C7.21"
7711..8433
/gene="SCO6485"
/notes="SC9C7.21, unknown, len: 240 aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein SC9C7.21"
/protein_id="CAA22733.1"
/db_xref="GI:4154080"
/db_xref="GOA:Q9ZBJ9"
/db_xref="SPTREMBL:Q9ZBJ9"
/translation="MSVTISRYDAWEGFNEASGDRGAVFWDABPTLTAGRHAFEP
HLTDAALPLIDLCGCTQTRYPAERPHVVGADLSAAALAHADPGGLAAVYRQLD
AADKAEQAALHAEGLDANVYVRGVLHOCEPDROPLADALAILGERGAEFLVRSQA
AKTVLMGLAQGPAPPAKAPVLRKHGLTPGAVPDEAVPEYLRAAGLTVIASGELPLVT
TEYTABDGTIELPSAWVAGTA"
8603..9436
/gene="SCO6486"
/notes="SC9C7.22"
/notes="synonym: SC9C7.22"
8603..9436
/gene="SCO6486"
/notes="SC9C7.22, possible transport associated protein,
len: 277aa; similar to SW:DPPA BACSU dppa, dipeptide
transport protein (274 aa) fasta scores; opt: 448,
z-score: 604.3, E(): 2.3e-26, (32.3% identity in 269 aa
overlap)".
/codon_start=1
/transl_table=11

```

Alignment Scores:

Pred. No.:	1.68e-65	Length:	296500
Score:	2493.00	Matches:	480
Percent Similarity:	67.33%	Conservative:	132
Best Local Similarity:	52.81%	Mismatches:	236
Query Match:	48.55%	Indels:	61
DB:	1	Gaps:	18

US-09-917-376-1 (1-957) x SCO939128 (1-296500)

QY - 10 ThrMetArgSerArgArgLeuValSerLeuLeuAlaAlaThrAlaSerPheAlaValAla 29

Db	70562	ACCGTGCGAAGAACCCGGATCCTCAACGTGCTGCTGGCCCTCGCGCGCGGTGTGTGGCG	70503
Qy	30	AlaAlaLeuGlyValLeuProIleAlaIleThrAlaSerProAla-----HisAlaAla	47
Db	70502	-----GGCAGTCCGCCCGGGCTGCGCGCGCGAGCCGCGCGCGGGCGGCTTGTCTCCCGGC	70452
Qy	48	ThrThrGlnProTyrThrTrpSerAsnValAlaIleGlyGlyGlyGlyPheValAspGly	67
Db	70451	GCCGCCGACAGCTACACCTGGAGAAGCCGCCATCGACGGCGGCGTTCGTCTCCCGGC	70392
Qy	68	IleValPheAsnGluClyAlaProGlyIleLeuTyrValArgThrAspIleGlyMet	87
Db	70391	ATCGTCTTCAACCGGACCGAGAAGCACTTGGCTTACGCCCGCACCGACATCGGAGCGCC	70332
Qy	88	TyrArgTrpAspAlaAlaAsnGlyArgTrrIleProLeuLeuAspTrpValGlyTrpAsn	107
Db	70331	TACCGTGGCAGGAGGTCCACACCTGGACCGCGCTCTCTGCACACGTGGCTGGGAC	70272
Qy	108	AsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysVal	127
Db	70271	GACTGGGGGCACACGGCGTGTGTCTCTCGCTCCGACCGCTCGATCCGACGGGTG	70212
Qy	128	TrpAlaAlaValClyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArg	147
Db	70211	TACGCGCGCGTGGCAGCTACACCAACGACTGGGACCGCACCAACGCGCGGTGTGGCG	70152
Qy	148	SerSerAspGlnGlyAlaValThrGlnIleThrProLeuProPheLysLeuGlyLysAsn	167
Db	70151	TCCGCCGACCGGGTTCGAGCTGGGAAGCGGACCTGCGCGTTCAGCTGGCGGGCAAC	70092
Qy	168	MetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeu	187
Db	70091	ATCCCGCGCGGGCATGGTGAGCGCTGGCGGTGACCCGCGACGACCAACGACGTGTG	70032
Qy	188	TyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrp	207
Db	70031	TACCTGGCGGCCCCAGCGGCACCGACTGTGGCGGTCCACCGACGGCGGTCACTGG	69972
Qy	208	SerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThr	227
Db	69971	TCCGAGGTACCGGCTTCGCCGAACCCCGGAACCTACGCGGAGACCCGAAACGACACTG	69912
Qy	228	GlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSer---Ser	246
Db	69911	GGTAGCGCTCGACAAACAGGGCATCACCCTGGTTCACCTTCGACGAGTTCAGCGGGCG	69852
Qy	247	SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProVal	266
Db	69851	GGCGCGGCACGCCACGCGGACCTCTACGTGGGGTCCGGGCAACAGAGAAACGCCGTC	69792
Qy	267	PheTrpSerArgAspClyGlyAlaThrTrpGlnAlaValProGlyAlaProThrClyPhe	286
Db	69791	TACCGTTCGACGACGCGGGCGGACCTGGGAGCGGCTCGCGGGACGCCGACGGGTTC	69732
Qy	287	IleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsn	306
Db	69731	CTGGCCCAACAGGTGTGTGGACGGGAGAACGGCTACCTGTACTTCGCTTACAGCGAC	69672
Qy	307	ThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGly	326
Db	69671	ACCGCGGCGCGTACGACGCGCGCAAGGGCGGCTGTACCGGTACCGCACGCGCACCGGC	69612
Qy	327	ThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheClyTyr	346
Db	69611	ACCTGACGGACATCAGCCCGCGCGGAGCGGACAC-----TACTACGGCTTC	69561
Qy	347	SerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSer	366
Db	69560	AGCGGCTGACCGTCGACGGCGGCGCGGAGCGGTGATGGCGACCGCTACAGCTCG	69501
Qy	367	TrpTrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIle	386

Db 176223 TGTTCACAGGTACACGACTTCCCGAAGCTCGGCACCTACGTGACGAGCGGACCGACAG 176282
Qy 227 ThrGlyTyrGlnSerAspIleGlnGlyValValTyrValAlaPheAspLysSerSer 246
Db 176283 AGCGGTACGCGTCCGACAAACAGGCGATCGTGGGTACCTTTCGACGAGTCGACGGG 176342
Qy 247 SerLeuGlyClnAlaSerLysThrIlePheValGlyValAlaAspProAlaVal 266
Db 176343 TCGCCGGGAGCTCCACGCGGAGCGGTACGTTCGGGTTCGCCCAAGACAACTCCGTC 176402
Qy 267 PheTyrSerArgAspGlyAlaThrTyrGlnAlaValProGlyAlaProThrPhe 286
Db 176403 TATTCGTTCACGACGCGGCGGACTGTGTCGCCGTGGCCCGACCGCCACCGGCGCAT 176462
Qy 287 IleProHisLysGlyValPheAspProValAlaHisValLeuTyrIleAlaThrSerAsn 306
Db 176463 CTCGCCCAACAGGGCGGTGTCGACGCGGCGAAGCGGTGTCGTACCTTCGGGTACAGGAC 176522
Qy 307 ThrGlyGlyProTyrAspGlySerSerGlyAspValTyrLysPheSerValThrSerGly 326
Db 176523 AAGGGCGGACCGTACGACGCGGCGCAAGGACAGCTGTGGCGGTACACGACGAAGACCGGG 176582
Qy 327 ThrTyrThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyr 346
Db 176583 ACCTGACGAAACATCAGCCCGGTTCGGGAGCGCGACAC-----TACTACGGCTTC 176633
Qy 347 SerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSer 366
Db 176634 AGCGGCTGACCTGACGCGGAGCATCCGGGACGGGTGATGGCGACTGGTACAGCTCC 176693
Qy 367 TrpTyrProAspThrIlePheArgSerThrAspGlyAlaThrTyrTrpThrArgIle 386
Db 176694 TGTGTCGCGACACGCGACTTTCGCTCCACGACAGCGGCGGACCTTCGACGAAGGCC 176753
Qy 387 TrpAspTyrThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGlu 406
Db 176754 TGGGACTACACCTCGTATCCGAGCGCTCGAAGCGCTTCACCATGTGTCTCGTCCCTCG 176813
Qy 407 ProTyrLeuThrPheGlyValGlnProAsnProProValProSerProLysLeuGlyTyr 426
Db 176814 CCTGTGCTACCTTGGGAGGAAACCCCGCACCCCGGACGAGACCCCGAAGCTCGGCTGG 176873
Qy 427 MetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGly 446
Db 176874 ATGACCGAGTCCCTGGAGATCGACCCGCTTCGACTCCGCGCGCATGTATGACGAACGGG 176933
Qy 447 AlaThrLeuTyrAlaThrAsnAspLeuThrLysTyrAspSerGlyGlyGlnIleHisIle 466
Db 176934 GCGACGGTCTACGCGCACGACCACTGACCAACTGGGACAGCGGAAGCCAGTTTCAACCATC 176993
Qy 467 AlaProMetValLysGlyLeuGluThrAlaValAsnAspLeuIleSerProProSer 486
Db 176994 AAGCCGATGGCGGGGCTTGGAGGAGACGGCGGTCAACGCTCGCCTCGCCTCCCTCC 177053
Qy 487 ---GlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyPheThrHisAlaAspVal 505
Db 177054 GCGCGGCCGACGCTGTTCAGCGGCTCGGTGACATCGCGGCTTCGGGACACGAGCTC 177113
Qy 506 ThrAlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSerValAsp 525
Db 177114 ACCACGGTGGCGTGGTGTATGATACGTCGCCGAACTTACACGACGACGACGCTCGAC 177173
Qy 526 TyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSerGln 545
Db 177174 TACCGCGAGACCAACCGGCGACGGTGGTGGCGGTTCGGCAATCTCGAC-----TCGGGT 177227
Qy 546 ProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTyrPheGlnGlySer 565
Db 177228 CCG-----CATGTGGGCTTTCGACGACCAACGCGGCCAATCTCGTTCGGGGGCGG 177278
Qy 566 GluProGlyGlyValThrThrGlyThrValAlaAlaSerAlaAspGlySerArgPhe 585
Db 177279 GACCCCTTCGGGGGTTCAGCGGGGTGGGACGCTCGCGGCGCGTCCGACGCGGCTTC 177338

Qy 586 ValTyrAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSer 605
Db 177339 GTGTGAGCCGCGCGGACCGGG-----GTGCAGTACACGACCGGGTTCGGCACTCG 177392
Qy 606 TrpAlaAlaSerGlnGlyValProAlaAlaAsnIleArgSerAspArgValAsnPro 625
Db 177393 TGTTCGCGCTCCGCGGCGCTCCCGCGCGGCGCATCTCGATCGGACCGGGTTCACCGG 177452
Qy 626 LysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThr 645
Db 177453 AAGACCTTCTACGGCTTCAAGTCGCGCAGGTTCTACGTCACTTCGACGCGGGGCGGAC 177512
Qy 646 PheGlnProValAlaAla---GlyLeuProSerSerGlyAlaValGlyValMetPheHis 664
Db 177513 TTCACGCGCTCGCGGCGCACCGGCTTCGCGAGCGCGACAGC-----GTGCGCTTCAAG 177566
Qy 665 AlaValProGlyLysGlyGluGlyAspLeuTyrLeuAlaAlaSerSer----- 679
Db 177567 GCGCTGCGCGCACGAAGGCGCATCTGCTGCGCGCGCGGCGGACGCGCGCTAC 177626
Qy 680 GlyLeuTyrHisSerThrAsnGlyLysSerTyrSerAlaIleThrGlyValSerSer 699
Db 177627 GGGCTGTGGCACTCGACGAGCGCGCGCGCTTCCAAAGCTCCGACCGTCCACAG 177686
Qy 700 AlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheVal 719
Db 177687 GCGGACACCATCGGCTTCGCAAGCGCGGCGCGCTCTGTACGACGCTCTACAC 177746
Qy 720 ValGlyThrIleGlyGlyValThrGlyAlaTyrArgSerAspCysGlyThrThrTrp 739
Db 177747 AGCGCAAGATCGCGGTGTGCGCGCATCTTCGCGTCGACCGCACAGGCGCGAGCTGG 177806
Qy 740 ValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTyrGlyGlnAlaIleThrGlyAsp 759
Db 177807 ACCGCGTCAACGACGATGCCACAGTGGGTGGAGCGGCGCGGCGCATCACCCTGAC 177866
Qy 760 HisAlaAsnLeuArgArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAsp 779
Db 177867 CCCAGGCTCTACGGCGCGGTGTATGTGTGACGAACGGCGCGGATCGTCTACGCGAC 177926
Qy 780 IleGlyAlaProSerGly-----SerProSerProSerValSerProSerAla 796
Db 177927 ACCGCGGCTCTTCGACGCGCGGCGGTACGAGCGCGGCGCACCGCGGCGCTGACGCGTG 177986
Qy 797 SerProSerLeuSer---ProSerProSerProSerProSerProSerProSer 815
Db 177987 ACATACAGGATCACGAACCAAGTGTGCGGCGGCTTCCAGGCGCGATGTACAGCTCCCAAC 178046
Qy 815 rProSerSerSerProSerSer-----Pr 824
Db 178047 ACCGGTTCGACCGCTGGGACGCTGCTGCTGGGCTGGTCTGTCGGGCGGAGGAG 178106
Qy 824 oSerProSer----- 827
Db 178107 GTACCCAGCTGTGGAACGCTCTGTACGCGCAGCGCGGTTTCGGGGGTGACCGCAGCGAAC 178166
Qy 828 -----ProSerPro----- 830
Db 178167 CTGGCTTGGAAACGGGAGGTTGGCGCGCGGCTCTCGGTGAGCTTCGGGTTCACGGGAAGC 178226
Qy 831 -----SerProSerProSerProSerProSerAla-----Se 843
Db 178227 TGTTCGGGATCCAAATGGCAGACCGCATCCGCTTCAAACTCGGCGGTGACAACTGACGCGTC 178286
Qy 843 rPro-----SerProSerSerProSerProSerSerProSerProSerProSer 861
Db 178287 GCCTGTAGCGAAGCGCA-CGCTCGCTGATCCGAGGCGCGCGCGCGCGCGCGCG 178345
Qy 861 rProThrProSerSerProValSerGlyValLysValGlnTyrLysAsnAsnAs 881
Db 178346 CCCTCCGCTCTCCACGCGCGTGTAGACCGTGGCGCTG-----GGCAA 178387

1156 GGGCCATATAATGGTGAATATGGAGAGATGATGAGATACAACTAAGACCGGAGANTGG 1215
329 ThrArgIleSerProValProSerThrAlaAsnAspTyrPheGlyTyrSerGly 348
1216 AAAAATAATAAGTCCATGGCTGCTCAAGATAC-----TATTTGGATATGGTGA 1266
349 LeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerThrPhe 368
1267 TTGGCTGTAGAGCCCGCAGAAAGTAGTATGATGATGATGATGATGATGATGATGATG 1326
369 ProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrThrPheArgIleThrAsp 388
1327 CCAGATACCTATATTTGGAGAGATGATGATGATGATGATGATGATGATGATGATG 1386
389 TrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrp 408
1387 TGGATGGCTATCCTTAATAGAACATTTGCATTAAACATGATGATGATGATGATGATG 1446
409 LeuThrPheGly---ValGlnProAsnProValProSerProIleLeuGlyTyrMet 427
1447 CTAAACTTTTGGAAATCTATCTTACACCACTGAAAGTAAAGCCCAAACTTGGCTGGATG 1506
428 AspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAla 447
1507 GTTGGAAACACTTGAATAAGACCATTTAATTCAGATARAATGCTTTATGGAACAGAGCT 1566
448 ThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAla 467
1567 AATTTGTATGGATGCGATGATTTGACAAATGGGACAAAGGGCAAAATATCACCATTA 1626
468 ProMetValLysGlyLeuGluGluThrAlaValAsnAspLeuIleSerProSerGly 487
1627 GTCAGGCTATAGGAATTTGAAGAACTTCAGTCCAGCACTGATAGCCCAACAGTTGGA 1686
488 AlaProLeuIleSerAlaLeuGlyAspLeuGlyPheThrHisAlaAspValThrAla 507
1687 CCACATTTATTCAGTGCACCTGGGAGATATAGCGGATTCAGACATGAAGACTTTGGAGAA 1746
508 ValProSerThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyrAla 527
1747 GCACCTTAATTTGGACATATGTTTCAAGCTTATGAGGACTACTACTGATATGATTTTGT 1806
528 GluLeuAsnProSerIleValArgAlaGlySerPheAspProSerSerGlnProAsn 547
1807 GAGTTAAATCCAAATTTATGGTTGTTAGTGGAAATCTTGATAAACAAGTGGATCCAAAT 1866
548 AspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGly---SerGlu 566
1867 ACTAACAGAAATGGTTTCTCATATGATGATGATGATGATGATGATGATGATGATGAT 1926
567 ProGlyGlyValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheVal 586
1927 CCTCAAGGAACACGGAAGTGTGACAGTTGTCAGCTGCTGCTGATGATGATGATGATG 1986
587 TrpAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrp 606
1987 TGGGCAACCA-----AAAGAGCGAAGATGATGATGATGATGATGATGATGATGATG 2040
607 AlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLys 626
2041 GTTGAATGCGCAAAATGTACCTTCAGAGCTATGTTATATCTGACAGAGTAAATCCAAAT 2100
627 ThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPhe 646
2101 AAGTTTTATGCAATTAAGATGGCAAAATTTTATATAGTCTGCTGATTAAGGAAACATTT 2160
647 ---GlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAla 665
2161 ATCGAAGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2208
666 ValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSerThr 685

2209 GTTCCAGGTATAGAGGTGATATCTGGCTAGTTGGTATATATGGATCTGGCACTCAACT 2268
686 AsnGlyGlySerSerTrpSerAlaIleThrThrGlyValSerSerAlaValAsnValGlyPhe 705
2269 GATGGTGGATATTCATTTGTAATAAATTTTCAAGTGTGAGGATGCGAGCGAGTATAGATTT 2328
706 GlyLysSerAlaProGlySerSerTyrProAlaValPheValValGlyThrIleGlyGly 725
2329 GGTAACCTCTGTAAGGAGAAACATATCTCTCAATTTATCTATGTCAAAAATTAATGGA 2388
726 ValThrGlyAlaTyrArgSerAspAspCysGlyThrThrTrpValLeuIleAsnAspAsp 745
2389 GTAAGAGTATATTTAGATCTGATGATGATGATGATGATGATGATGATGATGATGATG 2448
746 GlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgArg 765
2449 AAACATCAGTTTGGTCTGCGAATGCAGACATAAAGTGGGCTGAAATTTGCT-----TAT 2559
766 ValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGlyGlyAlaProSer 785
2509 GTATTTTGTGCAACAAACGGCTGGGTATAAATGGGGTGAATTTGCT-----TAT 2559
786 GlySerProSerProSerValSerProSerAlaSerProSerLeuSerProSerProSer 805
2560 TCAATATTTTGGCTATCAGCACTCCAACTCAACATCAACCAAGTTTCCCCGACCGGCT 2619
806 ProSerSerProSerProSerProSerProSerProSerProSerProSerProSerProSer 825
2620 CCAACGCAACTATTATCATCCACCTGCACCAACATCAACCAACCAACCAACCAACCAAC 2679
826 ProSerProSerProSerProSerProSerProSerProSerProSerProSerProSerProSer 845
2680 CCAACCAACCAACGAGCAACGCAACGCGCACGCGCGCTTACAGTACGCGTACT 2739
846 ProSerSerProSerProSerProSerProSerProSerProSerProSerProSerProSer 865
2740 ATGATCCGACCCGACCA-----ACACCAACCGCGCACCAACA 2781
866 SerSerProValSerGly---GlyValLysValGlnTyrLysAsnAsnAspSerAlaPro 884
2782 GGGACACTGTGACGGGAGTGGTTTGAAGTACTGTACAGAACATCAGACGAGTGGC 2841
885 GlyAspAsnGlnIleLysProGlyLeuGlnValValAsnThrGlySerSerSerValAsp 904
2842 AGTCAGGTTCTATAAGCGCTGGTTAAAGATAGTGAATGGAGGCGAGCAGCAGTGTGAT 2901
905 LeuSerThrValThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuVal 924
2902 CTTAGCAGGTTAAGATAAGATCTGTTACAGTGGATGGTGACAGCCAGAGTGGC 2961
925 TyrAsnCysAspTrpAlaAlaIleGlyCysGlyAsnIleArgAlaSerPheGlySerVal 944
2962 GTA---TGTGACTGGGACAGATAGTGGCAAGCATGTGACATTCATTTTGTGAAGCTT 3018
945 AsnProAlaThrProThrAlaAspThrTyrLeuGln 956
3019 AGCAGCGAGTGGAGTGGAGCGGATTTATTACTTGGAG 3054

RESULT 5

AX565635
LOCUS 2823 bp DNA linear PAT 29-NOV-2002
DEFINITION Sequence 1 from Patent WO02077242.
ACCESSION AX565635
VERSION AX565635.1 GI:26000969
KEYWORDS
SOURCE unidentified
ORGANISM unclassified.
REFERENCE 1
AUTHORS Duffner, F. and Sjoeholm, C.
TITLE Family 74 xyloglucanases
JOURNAL Patent: WO 02077242-A 1 03-OCT-2002;
Novozymes A/S (DK)


```

FEATURES
  source      Location/Qualifiers
1..2823
  /organism="unidentified"
  /mol_type="unassigned DNA"
  /db_xref="taxon:32644"
  /notes="Jonesia sp., DSM 14140"
1..2823
  /notes="unnamed protein product"
  /codon_start=1
  /protein_id="CAD58415.1"
  /db_xref="GI:26000970"
  /db_xref="RENTREMBL:CAD58415"
  /translation="MLRITVAVALLAALLATTAGGAATASTPMPTVAPSPFSSPSGSGY
  SWSNIVGGGFPVGFVFNQKDFGLVYARDTGAYRLANDSTGRWPLPLTDHIGWDDMS
  HSGIILSLADPVDVTRVLAAGTYSMDDPQNGAILRSADKGETWKTMLPFRVGNM
  PKRGMERLAVDPNNKVLVFGAESNGLWKSVDYKWTGKVTSTFNAGNYVADASGA
  YTGQNGVWFDPDTSKAGQGTQFLYGVADKQNNVSTDDGGATWQRPQPTGPF
  LAQGVLDHKGQQLYIATSDTGPYDGSKGVMDLSSQWTRISPISTSNAPF
  YSGLAIDRKNPDTIMVQVSWMPDMYVYRSTDRGTWSPIMELNGSQRTKQYNDY
  SGAPWLDFTAKPEANPKLGMWTSFIDPHNSDRFFYTGAGIYGGTNTLWNDKG
  KKVDTITVAQGIETAAQDLAAPPGNIGLYSALADIGFTKDIISOVPNKYVKNPHH
  DVTVTSIDFAESKPATVRAKKSISGRTTSMVGVSTDAKGTWPGATPSGVKPGSLTV
  SANASIVNAPEGAAPRRSTNGSSWTSVGLPHNAQVADSDRVNNTLIFGVDCKYH
  STNGGASFTASAPFTPTSGNVRFRAVPQKQHLWLAGGVSGSTGYGMRSTDDGKNWT
  KVGVOGDAVGFGKATSSGSPVIFTSAKIDGVRIFRSDDGKTKWIRINDNHNQWY
  WTGASITGDPDVEGRVYITNGRGIIVGDSSTPPPGQDSGASGNTGATPTPETDS
  DINTGSGTDTGTEBPGTSGSCTAPDSGNTSPGNTSSCKVYSTTDMGSGFTGATIT
  TMTSPSPINGWTALFTYPSGQTISSVWSATQTLISGRNVLKNSGWNPITPAGSVFEG
  FNAGYTSKNTPTTSPFSLNGVTCALAD"
1..69
  /sig_peptide
70..2820
  /mat_peptide
  /product="unnamed"

ORIGIN
Alignment Scores:
Pred. No.:      8,836-58      Length:      2823
Score:          2174.00      Matches:      430
Percent Similarity: 59.48%      Conservative: 141
Best Local Similarity: 44.79%      Mismatches: 311
Query Match:      42.34%      Indels:      78
DB:               6          Gaps:       16

US-09-917-376-1 (1-957) x AX565635 (1-2823)
QY 10 ThrMetArgSerArgArgIleValSerLeuLeuAlaAlaThrAlaSerPheAlaValAla 29
Db 13 ACCGTGGTGGCCACAGCACTAACTGCGCTGCTCGCCACGACTGCGCGTGGCATCGCGGCC 72
QY 30 AlaAlaLeuGlyValLeuProIleAlaIleThrAlaSerProAlaHisAlaAlaThrThr 49
Db 73 ACATCA-----ACCCCATGCCAGTAACTGACCGCCCTCTTCTCATCATCACT 123
QY 50 GlnProTyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleVal 69
Db 124 TCCGGCTATAGCTGGTTCGAACGTTGAGATTGTGGCGGGGCTTTTGTGCTCGAATTGTC 183
QY 70 PheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyMetTyrArg 89
Db 184 TTTTAACCAAGAACAGCCAGCTGGTGTATGCTGCGGCACACAGATATGCTGGAGCGTACCGG 243
QY 90 TrpAspAlaAlaAsnGlyArgTTrpIleProLeuLeuAspTrpValGlyTTrpAsnAsnTrp 109
Db 244 CTCATGACTCCACCGGGGCTTGGATCCCTCTTACCGACCATATATGTTGGGATGACTGG 303
QY 110 GlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAla 129
Db 304 AGCCACTCCGGAATCTGCTCTGGCCACTGACCTCTTGGACACCAACCGCGGTACCTT 363
QY 130 AlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSer 149
Db 364 GCAGCAGCAGCAGCTACTCCATCGGACTGGGATCCACAAAACGGTCTATTTTTCGCTCCGCG 423
QY 150 AspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetPro 169

```

```

424 GACAAAGCGGAGACCTGGGAAAAGACGATGCTGCCGTTCGGTGGGAGAACATGCCG 483
QY 170 GlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPhe 189
Db 484 GGTGCGGGATGGGAGAACGCTCGCCGTTGACCCAAACATAATAAAGATCTCTACTTT 543
QY 190 GlyAlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGln 209
Db 544 GGTGCGGAATCGGGTAATGGACTATGAAGTCAACTGACTACGGGAAAAACCTGGGGCAA 603
QY 210 MetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGly--- 228
Db 604 GTGACGAGCTTCCCAATGCTGGTAACACTACGTCGCC-----GATGCTTCTGTGTCG 654
QY 229 TyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerLeu 248
Db 655 TACACCGGGCAAAACCAAGGTGTGTGTGGGTAACTTTTGACCCCATCAGCGAAAGCT 714
QY 249 GlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrp 268
Db 715 GGTGACACACACAAACGATCTACGTAGTGTAGCAGACAGCAGACAACTTATTCGA 774
QY 269 SerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIlePro 288
Db 775 TCTACCGATGGCGGGCAACATGCGCAACGCTCCCTGGACAAACCCACCGGATTCCTGGCG 834
QY 289 HisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGly 308
Db 835 CAAAAGGGAGTGTGTGACCAATAAAGGTCAGCAACTCTATCATCGCTACCTCAGATACTGGT 894
QY 309 GlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGlyThrTrp 328
Db 895 GGCCCGTATGACGGGTGAGAGGGGATGTGTGGCGCTAGATATTTCCAGTGGCCAGTGG 954
QY 329 ThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyrSerGly 348
Db 955 ACTCGAATCAGCCCAATTCCTCTTACC--TCATCGAATTCAGCGTTCGGGTACTCCGCT 1011
QY 349 LeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrpTrp 368
Db 1012 TTAGCCATTTGACCGCAAAACCCCTGACACCATCATGTTGTTTCCCAAGATCTCTTGGTGG 1071
QY 369 ProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIleTrpAsp 388
Db 1072 CCTGACATGATGATATTCGCTCCACCGCGGGGAAAAAGTGTGCTGCTATCTGGGAA 1131
QY 389 TrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrp 408
Db 1132 CTGAACCGTTCCACAGCCGGTACAAAGCAGTACACACGATTACTCAGGGGCTCCCTGG 1191
QY 409 LeuThrPheGlyValGlnProAsnProProValProSerProLysLeuGlyTrpMetAsp 428
Db 1192 CTTGACTTTGGCAATACAGCCAAAGAGCCTGAAGCTAACCCCAAACTCGGGTGGATGACC 1251
QY 429 GluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThr 448
Db 1252 CAGTCTTCCGAGATTGACCCGCATAACTCCACCGCTTCTTCTATGGCACCGCGCGGGC 1311
QY 449 LeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAlaPro 468
Db 1312 ATCTATGGCGGAACCACTCACAACTGGGATAAGGGCAAAAGATTGATATCACTGTC 1371
QY 469 MetValLysGlyLeuGluThrAlaValAsnAspLeuIleSerProSerGlyAla 488
Db 1372 AAGGCCCAAGGATCGAAGAAACCGCAGCCCAAGACCTTGGCGCTCCCGCTCGAAACATT 1431
QY 489 ProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAlaVal 508
Db 1432 GGCCTATACTACGCGCTGGCAGACATCGGTGGCTTCCACCAAAAGATATCTCTCAGGGT 1491
QY 509 ProSerThr---IlePheThrSerProValPheThrThrGlyThrSerValAspTyrAla 527

```

Db 1492 CCTAACAGTACTACTACAGAACCCACACCATGACACCGTGACCTCCATGACTTCGCC 1551
 Qy 528 GluLeuAsnProSerIleValArgAlaGlySerPheAspProSerSerGlnProAsn 547
 Db 1552 GAAAGCAAAACCGCAACTGTCGTGCGTGCAGGCAAA-----TCTATTTCTGGGGAACC 1605
 Qy 548 AspArgHisValAlaPheSerThrAspGlyGlyIleAsnTrpPheGlnGlySerGluPro 567
 Db 1606 ACCTCATGGGTTTCACTGATGCTGCGGAAACGCGGAAACCTGCGGCAACGCCA 1665
 Qy 568 GlyGlyValThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTyr 587
 Db 1666 TCGGGGTTAAAGCCCGCTCCATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1725
 Qy 588 AlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAla 607
 Db 1726 GCCCGG-----GAAGTGCGGCTCTCGACGTTCCACCAACAGTGCGCTCTCGTGCTCA 1779
 Qy 608 AlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThr 627
 Db 1780 ACGGTGAGCGGCTCTCCATGACGAGTGCATGCGCATCAGACCGAGTCAACGCTAATPACT 1839
 Qy 628 PheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGln 647
 Db 1840 CTTTACGGTTCCTGACGGAAGTTCTATCATCTCCACAAATGGTGGGCTTCTCTCAGC 1899
 Qy 648 ProValAla---AlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaVal 666
 Db 1900 CGCTCAGCTTTCACGGGTTTCCACACGATGGA-----AAGCTTCGTTCCGTCGCGTG 1953
 Qy 667 ProGlyLysGluGlyAspLeuTrpLeuAla-----AlaSerSerGlyLeu 681
 Db 1954 CCGGGCGCTCAGGTCATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2013
 Qy 682 TyrHisSerThrAsnGlySerSerTrpSerAlaIleThrGlyValSerSerAlaVal 701
 Db 2014 TGGCGGTCCACTGATCGCGGGAAGAACTGACCAAGGTATCTGGAGTTTCAGGAAGCGCAT 2073
 Qy 702 AsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheValValGly 721
 Db 2074 GCGGTGGTTTCGTAAGCAACCAAGTAGTAGGATACCGAGTATTTTACATTCGCC 2133
 Qy 722 ThrIleGlyValThrGlyAlaTyrArgSerAspAspCysGlyThrTrpValLeu 741
 Db 2134 AAGATTGATGGAGTTCGAGGAATTTCCGCTCTGACGATGAGCAAGACCTGGAACGC 2193
 Qy 742 IleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAla 761
 Db 2194 ATCAACGACACACCAACCAACGATGATGAGTGGCGCATCAATCACCCTGACCCAGAT 2253
 Qy 762 AsnLeuArgArgValTyrIleGlyThrAsnGlyValGlyValTyrGlyAspIleGly 781
 Db 2254 GTCAAGGGCGGCTTACATTTGATGATGATGATGATGATGATGATGATGATGATGATG 2307
 Qy 782 GlyAlaProSerGlySerProSerProSerValSerProSerAlaSerProSerLeuSer 801
 Db 2307 ----- 2307
 Qy 802 ProSerProSerProSerSerProSerProSerProSerProSerProSerProSer 821
 Db 2308 -----TCCTCCATCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCATCG 2352
 Qy 822 SerSerProSerProSerProSerProSerProSerProSerProSerProSerPro 840
 Db 2353 GGAATACAGGGGCCACACCGCTGAAACCCCTGACAGCGATCAACACCTGCGCGG 2412
 Qy 841 SerAlaSerProSerProSerProSerProSerProSerProSerProSerProSer 860
 Db 2413 TCCACTGATACGGCGGCAACTTACAGCCCGGGAACAGGAGTGGATCGTGGACCGCACT 2472
 Qy 861 SerProThrProSerSerProValSerGlyGlyValIleValGlyValGlyValGly 880
 Db 2473 GACTCGGGGAACACCTCTCTCGGGAATACCACTGATGATGATGATGATGATGATGATG 2532

Qy 881 AspSerAlaProGlyAspAsnGlnIleLysProGlyLeuGlnValValAsnThrGlySer 900
 Db 2533 GACTGGGGCAGCGC-----TTCACTGGGGCAGCGACCATCACTAACACGATCCCCC 2583
 Qy 901 SerSerValAspLeuSerThrValThrValArgTyr----- 912
 Db 2584 AGCCCAATTAACGGCTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2643
 Qy 913 ---TrpPheThrArgAsp-----GlyGlySerSerThrLeuValTyrAsnCysAspTrp 929
 Db 2644 GTGTGGTTCAGCAACTCAGACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2703
 Qy 930 AlaAlaIleGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 949
 Db 2704 -----AATCCAACGATACCT 2718
 RESULT 6
 AE007608 12732 bp DNA linear BCT 27-JUL-2001
 LOCUS Clostridium acetobutylicum ATCC824 section 96 of 356 of the
 DEFINITION complete genome.
 ACCESSION AE007608 AE001437
 VERSION AE007608.1 GI:15023819
 KEYWORDS
 SOURCE Clostridium acetobutylicum
 ORGANISM Clostridium acetobutylicum
 Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 Clostridium.
 REFERENCE 1 (bases 1 to 12732)
 Nolling, J., Breton, G., Omelchenko, M.V., Markarova, K.S., Zeng, Q.,
 Gibson, R., Lee, H.M., Dubois, J., Qiu, D., Hitti, J., Wolf, Y.I.,
 Tatusov, R.L., Sabathe, F., Doucette-Stamm, L., Soucaille, P.,
 Daly, M.J., Bennett, G.N., Koonin, E.V. and Smith, D.R.
 Genome sequence and comparative analysis of the solvent-producing
 bacterium Clostridium acetobutylicum
 J. Bacteriol. 183 (16), 4823-4838 (2001)
 21359325
 11466286
 REFERENCE 2 (bases 1 to 12732)
 Childress, D., Zeng, Q. and Smith, D.R.
 Direct Submission
 Submitted (24-JUL-2001) GTC Sequencing Center Production,
 Finishing, and Bioinformatics teams, Genome Therapeutics Corp., 100
 Beaver Street, Waltham, MA 02453-8443, USA
 FEATURES
 Location/Qualifiers
 source
 1..12732
 /organism="Clostridium acetobutylicum"
 /mol_type="genomic DNA"
 /strain="ATCC 824"
 /db_xref="ATCC:824"
 /db_xref="taxon:1488"
 73..2592
 /gene="CAC0919"
 73..2592
 /gene="CAC0919"
 /codon_start=1
 /transl_table=1
 /product="Probably secreted sialidase; several ASP-boxes
 and dockerin domain"
 /protein_id="AAK78895.1"
 /db_xref="GI:15023820"
 /translation="MNKRIVSMVAGLSITFTTFVTHISAAANKAAASVQSGYKWDN
 AKIGAGGVPAVIFNTEKDLIYARTDMGAYRWDKANNKIPIITDGSDFMTLGCES
 IATPIDITNRVITIAAGLTINDWQDENATILSSQDKGTNWKRLQLEPFKVGNNPGRNMG
 ERQIDIPNDNKILYLGARSGNLKSDYGTWSKVDNFPDPTGVDQDPNEYTADKV
 GVWETFDPSLTGSGTPTMYGAADTKNNIYVINDGKTWSAVKQPGKGLPHHG
 ILASDGLIYSISNTSGPDGSGQVKNYKTGWTNITPPAVGDTKSGFGISVDA
 ONPNVVATLNRWPDLELYRSTDAKTWKPIWNGYVNRITLGVNLDYSAQPLDW
 KGTGVTTPDLVPLKLGMMGLLEIIPFNSDRPFYGTGLYGLTDDTLNWDKGNVLSV
 KANGIECANVDVVPTGAQLSAGDDCGFYHDDITKVPKSMKMTTNFSAITSLDY
 AESVNFVFRVGVNVDTSKNQDQCGISYDGKKNWFSAGSNISGVYKAGTVAAGDAK
 TIVSPBEGANAAYSTDNKNTKWTCSGLPQGAQKVRSDRVNPKKFGYGLNGKFIYSTDA

gene
CDS
GATFTOSSQTLTKGKIGPKTVIHGEGDIWAGKXGDLWHSTDSGATFTKVSVDAS
DVTVLGSKIDDDGPAIYMDATIDGTAGIFRSDDEGATWVRINDDAHQVGSPICTIG
DPNKYGRVFGVINGRIVGSDQPTPTSPVTLGVDNDGIIINGRDVWVLTOYTAG
KQVTDIKAADVNGDGTNGRDLMTLRQLQISGRITSPSTK"
complement (2634. .2978)
/gene="CAC0920"
complement (2634. .2978)
/gene="CAC0920"
/codon_start=1
/transl_table=11
/product="Protein related to MIFH/DOPD protein family,
function in bacteria is unknown"
/protein_id="AAK78896.1"
/db_xref="GI:15023821"
/translation="MPEFINSTVTVKLDKEKSTLKVGLKIIIEILLPKSDDWLVGFK
DNPVYFKGKQKKAFAIEVKIGSADKASKNKLVAISSIMENELSPKDCIVITVD
EVSNNWAWNSLP"
3104. .4252
/gene="CAC0921"
3104. .4252
/gene="CAC0921"
/codon_start=1
/transl_table=11
/product="Metal-dependent amidohydrolase"
/protein_id="AAK78897.1"
/db_xref="GI:15023822"
/translation="MITIEQBEFLKLLIGYRELKHPLSKMEYETTKRIKRWLVE
NDIKVLDPTEVGVAIEIEGHEGKTVARADIDALPIEKTLPSSVNNKGIMHACG
HDPHTAIIIGTALILSKRAEIVGRVRIIPQGEETGKSGOYIEEGVLRDVCIFGM
HNPDPVGTIGSELMAEVDREIKVIGSGHAGIPNKCIDPIVVASQIVASQWOT
VVSRSIPESVITRFSNGTNNVSDSMEGTFTFNNIRERIEIRNLMERTAK
GIARALGKISYFVNSPVLNDVRFKOLCIAEADNGYVVEAKRNILGGEDPSFYQ
TVVPSFVWVGVDGTEEMHKANTYLNKEDALMVAANYPSDLALAKVLSK"
4594. .4863
/gene="CAC0922"
4594. .4863
/gene="CAC0922"
/codon_start=1
/transl_table=11
/product="Uncharacterized small conserved protein, homolog
of YfjA/yukE B. subtilis"
/protein_id="AAK78898.1"
/db_xref="GI:15023823"
/translation="MFDKAALCEDIKKSIEKSTTOLDSWYGESCTAFEKTVILDR
KLNTYTEILQEVSKALYDISEYQNTGDIKAKMSKAVNEASEQ"
5056. .6348
/gene="CAC0923"
5056. .6348
/gene="CAC0923"
/codon_start=1
/transl_table=11
/product="Predicted membrane protein; CF_1 family"
/protein_id="AAK78899.1"
/db_xref="GI:15023824"
/translation="MOSLKVDRNGLLKINEAKOGLKSAVDKPSLKNKRSRSTIELKN
LRELKAVNOLGEIRKINSNDYMKRFTDVKKCAKLTKSGTYTKKVLGNFGE
EVERPVIGIGVPIDAVESWDRNKFVLKTTITILETGAIGCVITLPEAGITWKA
WDLAAGVNSLDNIGTASAIYQRRKSEEDSSGYDLFQSPGGAGWSAKTLG
VKNPSYSCSGTXYDRSTAVLVNHSVTSGLGDKVEKLPALKKAESASANKTLGA
ANNAEKIHMADMMATKINGINNAKEVVDLKLKNVGTNFPYKQFYKDKNLL
KDGKGLPOLQDLLOQDLETFEKNITLNAKNYKNSIRDLNKNYRDIKASLYNIPA
GGTLNTEVDQGSFHNYPGEIKPIFEIKVVLGGER"
6345. .6923
/gene="CAC0924"
6345. .6923
/gene="CAC0924"
/codon_start=1
/transl_table=11
/product="Predicted membrane protein"
/protein_id="AAK78900.1"
/db_xref="GI:15023825"
/translation="MIRFOKRWAIFFLTISILIIICGDIKITSFGIEFDLILCNLPFI
GLGIIGFIYRWKQIKKYTELLKEDRLDKIEYIYDIERYVGVVYNTAFYIDGVK

gene
CDS
KLRPQYKCKDNNDYVLEPIYKXNGEFKDFIAYADKNDYKKSIESITNEITECEEK
AEKKRKRRLNEYVNSRMELAKKYEERAKNM"
7003. .8043
/gene="CAC0925"
7003. .8043
/gene="CAC0925"
/codon_start=1
/transl_table=11
/product="TPR-repeat-containing protein"
/protein_id="AAK78901.1"
/db_xref="GI:15023826"
/translation="MENENVLRQAQLNNRGLIFERNAGDEALKYFNKAIKEDENPKE
AYLNAALCLAMNVDDASCYDKLIIKYPNEGRAYPGKGNVFFKONIKRAIELYN
RSYLGSMSEGVYNLALCMESIGELDDAIKWLNRALINDQNIITLNKKAVALIKIK
KFTEALSCYRNLKIEVDNEVYHFKMLGEMGKIEKAFETISIGEALLGQMTILCY
DHALPFKGEFERALECTERALLPDEGNVLLIKKGLLLTVLKKIDEAKSYVDKILE
FHNMEGVFAKANILCMLLGEHEDAEFFKEIILRSGEESVLINSYVYRALNLKWA
KTRKQKRLIRKQ"
7998. .8357
/gene="CAC0926"
7998. .8357
/gene="CAC0926"
/codon_start=1
/transl_table=11
/product="TPR-repeat-containing protein"
/protein_id="AAK78902.1"
/db_xref="GI:15023827"
/translation="MGKNEEAKYKAEIKKYNELIKYPYDVKLNFPPKANCRLDIED
YKABEYIYIIDLNDALLEVYLMARNRISLNKYEAAKEDLNTVTKLNPAYRKVIEL
DEQLKEFLSGKSECYR"
8379. .8675
/gene="CAC0927"
8379. .8675
/gene="CAC0927"
/codon_start=1
/transl_table=11
/product="Hypothetical protein, CF-2 family"
/protein_id="AAK78903.1"
/db_xref="GI:15023828"
/translation="MNSDEIRIDEKACDDVVKDKAVGEKVKETYNLHLSNRHIEDG
IKGSVDSITTEYNKIVKMKEMQDNMGDVAANNIVSFSDAEDKMIKTQYR"
8713. .9012
/gene="CAC0928"
8713. .9012
/gene="CAC0928"
/codon_start=1
/transl_table=11
/product="Uncharacterized small conserved protein, homolog
of YfjA/yukE B. subtilis"
/protein_id="AAK78904.1"
/db_xref="GI:15023829"
/translation="MVDRCIDIEASIFETIKVYSDSKVRLHDLCDLENELKQMED
TWEQDAKDFDSTPPGFSAMKQCDMLLEELIKLVKTSFESLDTEMKNLDKX"
9712. .10335
/gene="CAC0929"
9712. .10335
/gene="CAC0929"
/codon_start=1
/transl_table=11
/product="SAM-dependent methyltransferase"
/protein_id="AAK78905.1"
/db_xref="GI:15023830"
/translation="MGSNQIFDNTAASNVIRSYFERLKYVLSKVISKDKITADP
GCGTFTISALASDANVFGLDNSNMLRELKSAHDKLNNIYLLKSLTSVTLFDE
SIDAVFTNMLHHIKDAEKAKENYRVLKQGTTVISDVSEHNGKAREEMFDELWG
SNEQIDMWLNKVGFKNIIDIENTDLKCKGTSSRGEHTSTSIKAAK"
10353. .11498
/gene="CAC0930"
/note="metB"
10353. .11498
/gene="CAC0930"
/codon_start=1
/transl_table=11

Alignment Scores:
 Pred. No.: 5.06e-53 Length: 12732
 Score: 2037.00 Matches: 392
 Percent Similarity: 64.39% Conservative: 218
 Best Local Similarity: 49.49% Mismatches: 152
 Query Match: 39.67% Indels: 30
 DB: 1 Gaps: 16

US-09-917-376-1 (1-957) x AB007608 (1-12732)

QY	13	SerArgArgLeuValSerLeuLeuAlaAa	-----	ThrAlaSerPheAla	27
DB	76	AATAAAAGGATGTAGCATGGTTGCAGCACTAGCATATATTTTCACAACTGGTTTT	---	---	132
QY	28	ValAlaAlaLeuGlyValLeuProIleAlaIleThrAlaSerProAlaHisAlaAla	47		
DB	133	-----GTAACACATATAAGTCGACGAGCTAATAAGCAGCAGCACTTCA	---	---	174
QY	48	.Thr---ThrGlnProTyrThrTrrpSerAsnValAlaIleGlyGlyGlyPheValaAsp	66		
DB	175	GTTCAAAGCCCAAGGATATAAGTGGGATTAATCAAAGATAGGTGCTGGTGATATGTTCCA	---	---	234
QY	67	GlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGly	86		
DB	235	CGAGTTATATTATAAACACAGAAAAAGATTTAATATGATCTCGTACAGATATGGAGGA	---	---	294
QY	87	MetTyrArgTrrpAspAlaAlaAsnGlyArgTrrpIleProLeuLeuAspTrrpValGlyTrrp	106		
DB	295	GCATATAGATGGGACAGGCTTAATAAATAGTGGATTCCAAATTTACTGAT-----GGATTT	---	---	348
QY	107	AsnAsnTrrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLys	126		
DB	349	TCTGATTGGACAATGCTTGGCTGGCAAGATATAGCACTGACCCAAATAGACACTAATCGT	---	---	408
QY	127	ValTrrpAlaAlaValGlyMetTyrThrAsnSerTrrpAspProAsnAspGlyAlaIleLeu	146		
DB	409	GTTTATATAGCAGCAGGTTTCTACACAAATGATTGGCAAGATGAAATGCATATATCTTT	---	---	468
QY	147	ArgSerSerAspGlnGlyAlaThrTrrpGlnIleThrProLeuProPheLysLeuGlyGly	166		
DB	469	AGTTTCACAGGATAAAGGTAAACACATGGAAAAAGATACCAACTTTCCATTTAAGTACGAGGT	---	---	528
QY	167	AsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIle	186		
DB	529	AATATGCCGGAAGAATATGGGAGAAAGACTTCAATAGATFCCCTAATGATGATATAAATA	---	---	588
QY	187	LeuTyrPheGlyAlaProSerGlyLysGlyLeuTrrpArgSerThrAspSerGlyAlaThr	206		
DB	589	CTTTTATCTTGGAGCTAGAAAGTGGTAATGGACTTTTGGAAAAAGTGAAGACTATCGACAAACA	---	---	648
QY	207	TrrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThr	226		
DB	649	TGGTCTAAGGTGTGATAATTTTCCCTGATACAGCGCACTATGTACAAGATCCTCAAAATGAA	---	---	708
QY	227	ThrGlyTyrGlnSerAspIleGlnGlyValValTrrpValAlaPheAspLysSerSerSer	246		
DB	709	-----TATACAGCAGATAAAGTTGGAGTAGTTTGGGAAACCTTTGATCCTTCAACAGGA	---	---	762
QY	247	SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAsp---ProAsnAsnPro	265		
DB	763	ACAAAAAGGAAGCCCAACTCAAACTATGTATGTGTGGTCAGCAGATGAAGACAGGAACAAT	---	---	822
QY	266	ValPheTrrpSerArgAspGlyGlyAlaThrTrrpGlnAlaValProGlyAlaProThrGly	285		
DB	823	ATTTACGTTACTAATATGATGGAGGAAGACATGAGTGCAGTTAAAGGACAGCCAAAAGGA	---	---	882
QY	286	PheIleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSer	305		
DB	883	TATCTTCCACATCATGGTATACTT---GCTTCCGATGGTATGCTTTATATATCATATAGC	---	---	939
QY	306	AsnThrGlyGlyProTyrAspGlySerSerGlyAspValTrrpLysPheSerValThrSer	325		


```

/db_xref="GI:9755897"
/db_xref="GOA:Q9P4T8"
/db_xref="SPTREMBL:Q9P4T8"
/translations="MPSRVKFWITLTLGTFSLVYQRAASQSYRWQNVKIGGGGFPVG
IVNPEKGLAYARTDIGAYKLNADDTWPLDPAADDSSRWNYGVDAALATDPVQNR
LYLATGLYNSWDPSNGHLLISSDYKGTDAAPLPEKLAGNMPGRGWERLVVDPNL
SLYFGARSGNGLKTSNGRSWKSUTFTDGSFVDPDSDTGLNSDKIGIAWVTID
KAGSGSATPRIFVGVANKSNSIYISTNGSSWSAVAGQPTSFILPHKGVLSPSBR
LYITYDGGGPYDTSGLVYKYSIDTAWNTDITPVSGLDFPFGGLALDITQXNGTVM
VAALNSWPDQOIFRSTNGCASWPLDWAIVYPLNKYYSNAALAPWIGPNVDVTP
GNLQIWMESLISIDPDSNMHLYGTETLYGSRDILLKWDSAHNVTIKSLADGVEITS
VQALISPPGSPPLVSAIDGVGFVHQSLETRAPSAQTNPKWTSADIDLRNPNPSVV
RIGTSGTSKQVSAISDGVGVWQHPADPNVGGKVAISADADILLRNTNGVNV
SRQATFNIFYGASRTFYSTDNGTFFSAHGLSGATSVDFTVHPVSVDGIWASDT
KGLFHTDSGATPSATSGITQAWGVALGAPRSTGYPAVFAAANYGGEVAYFRSDRG
VNWVKLNDAAHGFGAASACMAADPRVYGRVYIGTNGRGIIFYGDVAGSAPPPTSSATS
TTSVSITITRSTITRSTITRSTITRSTITRSTITRSTITRSTITRSTITRSTITRST
FISQCLPI"
sig_peptide      28..96
/gene="cel6"
mat_peptide      97..2445
/gene="cel6"
/product="CEL6"

```

ORIGIN

Alignment Scores:

```

Pred. No.:      4..9e-46      Length:      2579
Score:          1784.00      Matches:      355
Percent Similarity: 59.78%      Conservative: 131
Best Local Similarity: 43.67%      Mismatches: 277
Query Match:      34.74%      Indels:      50
DB:              8           Gaps:       13

```

US-09-917-376-1 (1-957) x ABI292929 (1-2579)

```

QY 20 LeuAlaAlaThrAlaSerPheAlaValAlaAlaLeuGlyValLeuProfilAlaile 39
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 28 ATGCCAGCAGAGTAAATTCGATAAATCTGCACACTTGGCACCCTTCAGCCTTGTCTAC 87
QY 40 ThrAlaSerProAlaHisAlaAlaThrThrGlnProTyrThrTrpSerAlaValAlaile 59
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 88 -----GGGCAAGCAGCAGCTTCTCAGTCTTAGGTGGCAGAAATGCAAGATC 132
QY 60 ---GlyGlyGlyGlyPheValaspGlyIleValPheAsnGluGlyAlaProGlyIleLeu 78
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 133 GCGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 192
QY 79 TyrValArgThrAspIleGlyGlyMetTyrArgTrpAspAlaAlaAlaSerGlyArgTrpIle 98
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 193 TATGCTCGCACTGACATCGGAGGAGCTACAGCTGAAT---GCTGATGATACTTGGACG 249
QY 99 ProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyrAsnGlyValValSerIleAla 118
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 250 CCTCTACTTGATTTGCTGATGATCTCGTGGTGAATTAITGGGGCGTGGATGCTCTTGA 309
QY 119 AlaAspProIleAsnThrAsnIleValTrpAlaAlaValGlyMetTyrThrAsnSerTrp 138
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 310 ACTGATCCCGTCAACCCACCGCTTTTACCTAGCGAGCTGGGTGTATACCAACTCGTGG 369
QY 139 AspProAsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThr 158
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 370 GACCAAGTAACGGCCATATCTCATATCAAGTGATATGTTAAACATTCGATCGCGCG 429
QY 159 ProLeuProPheIleLeuGlyAlaMetProGlyArgGlyMetGlyGluArgLeuAla 178
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 430 CCTTTGCCATTCAGCTCGTGTATATGCTTGGCGGTGAATGGAGAGCGTCTCGTC 489
QY 179 ValAspProAsnAsnAspIleLeuTyrPheGlyAlaProSerGlyIleGlyLeuTrp 198
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 490 GTGATCCCAATCTGAACAGCATCTTTATTTTGGCGCTCGCAGTGGTAAACGCTCTCTGG 549
QY 199 ArgSerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThr 218
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::

```

```

Db 550 AAGTCAACAAATTCGGTGGAGCTGGAGCAAGGTAAACAAGCTTCACTGACACTGGGTCT 609
QY 219 TyrIleAlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTrp 238
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 610 TTCGTTCCCGATCCATCAGATAGCACCGSCCTGAATTTCTGATAAGATTGGGATCGCATGS 669
QY 239 ValAlaPheAspLysSerSerSerSerLeuGlyGlnAlaSerLysThrIlePheValGly 258
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 670 GTTACCATCGATAAAGCGTCAAGGCTCTCTGGTTCGGCTACTCCCGTATATTTGTGGC 729
QY 259 ValAlaAspPro---AsnAsnProValPheTrpSerArgAspGlyGlyAlaThrTrpGln 277
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 730 GTTCGAAACAGGCTCAACAGCATTTATATATCTACGAATGGCGGTCTCTTTGGAGT 789
QY 278 AlaValProGlyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsn 297
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 790 GCGTCCGAGGTCAACCACTAGCTTCTGTCACATAAGGGCGTCTATCCCTTCTGAG 849
QY 298 HisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAsp 317
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 850 AGGCACCTTTATATACCTTACTCCGATGGTGGCGTCTTACGATGGCACAAAGCGTGCA 909
QY 318 ValTrpLysPheSerValThrSerGlyThrTrpThrArgIleSerProValProSerThr 337
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 910 GTTTACAAGTACAGCATTTGACATGGCGCATGGAGTGCATACATAACCCGTCAGCGGAAGC 969
QY 338 AspThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsn 357
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 970 GATCTC-----TTCCTTTGGTTTGGAGGTCTAGCCTTAGATACAGAAAAACGGG 1020
QY 358 ThrIleMetValAlaThrGlnIleSerTrpTrpProAspThrIleIlePheArgSerThr 377
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1021 ACTGTCATGTTGAGCCCTCAACTCGTGGTGGCTGATGTCAGATATTCGCTCGACC 1080
QY 378 AspGlyGlyAlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeu 397
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1081 AAGGGGGTGCATCTTGGACTCCGCTGGGACTGGGAGTCTACCCGACTCTCAACAAA 1140
QY 398 ArgTyrValLeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnPro 417
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1141 TACTACAGTTATTAACGCTGGCCCTGGCGCGTGG-----ATAGGACCCCAATC 1188
QY 418 ProValProSerPro-----LysLeuGlyTrpMetAspGluAlaMetAlaIleAsp 434
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1189 GTTGATGTCACCCCGGAAATCTACAAATCGGTGGATGATGATGATGATGATGATGAT 1248
QY 435 ProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAsp 454
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1249 CCTTTCGACTCGAATCATTTGCTCTAGCGCACTGGAGAACTATCTATGGAAGTCTGAC 1308
QY 455 LeuThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGlu 474
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1309 CTCTCAAAATGGGACAGTGCACACACAGCTCACACCTCAAGTCTGCTCGAGATGGTCTGAG 1368
QY 475 GluThrAlaValAlaAsnAspLeuIleSerProProSerGlyAlaProLeuIleSerAlaLeu 494
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1369 GAAACCTTCGCTCCAGCCCTTATTTTACCTCTTCGAGACCGCCCTCTTGTCTCTAT 1428
QY 495 GlyAspLeuGlyGlyPheThrHisAlaAspValThrAlaValProSerThrIlePheThr 514
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1429 GGTGAGCTTGGCGGTTTTGTCCCAATCACTTACCCGTCGACCCAGCGCCAGTTTACC 1488
QY 515 SerProValPheThrThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIle 534
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1489 AATCCCAAGTGGTGCAGCTCGGCTGACATCGATTTCGCTGGCAATAACCTAGTAAATGTC 1548
QY 535 ValArgAlaGlySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSer 554
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1549 GTTCGATTCGATTCGTGTAGTAGTAGTAGT-----GGCAACACAGGTCGCCATCTCT 1599
QY 555 ThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGlyValThrThrGlyGly 574
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1600 TCGACTATCGCGTGTACTTGAATCAGANTTTTGGAGTCTCTGATAATGTTCAAGCGGA 1659

```


425 GGGNATATGCGCGTCCGCGATGGGTGAGAGATTGGCGGTGCGACCGAACAAGATAGC 484
Db
186 IleLeuTyrPheGlyAlaProSerGlyLeuTyrArgSerThrAspSerGlyAla 205
Qy
485 ATTCTCTACTTTGGAGCGCGAGTGGGCACGCGCTTGTGGAAGAGACACAGATTATGAGCG 544
Db
206 ThrTrpSerGlnMetThrAsnProAspValGlyThrValIleAlaAsnProThrAsp 225
Qy
545 ACCTGGTCGAATGATCACTCTTCCACGTGAGTGAACAGTATTTCAGGATTTCGAGCTCG 604
Db
226 ThrThrGlyTyrGlnSerAspIleGlnGlyValIleValTrpValAlaPheAspLysSer 245
Qy
605 ACC-----TACACTTCGACCCCTGGGTATCTCTGGGTGACCTTTGATTCGACCTCT 658
Db
246 SerSerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnPro 265
Qy
659 GGTAGCTCGGGTCTGCGACCGCTAGGATCTTCTGGAGTCGGGATGCTGGCAAGTCG 718
Db
266 ValPheTrpSerArgAspGlyAlaThrTrpGlnAlaValProGlyAlaProThr--- 284
Qy
719 GTCTTCAAGTCGGAGGATGCGGGTGTCTACTTGGCTTGGGTGTCTGGAGAGCCTCAGTAT 778
Db
285 GlyPheIleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThr 304
Qy
779 GGTCTTCTGCGCACAGGGGGTCTTGTCCCTGAGGAGAGACGCTGTATCTCTAC 838
Db
305 SerAsnThrGlyGlyProTyrAspGlySerGlyAspValTrpLysPheSerValThr 324
Qy
839 GCGNATGGCGTGTCTTACGATGGACAAACGGAACAGTCCCAAGTACACATCACC 898
Db
325 SerGlyThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPhe 344
Qy
899 AGCGGATGTGGACGGATATAAGTCCC-----ACCTCCCTTGGAGTACCTACTAC 949
Db
345 GlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGln 364
Qy
950 GGCTACGGTGTCTGCTCGGTGACCTCAAGTTCCTGGACCTTGTATGGTGGCTGCACTC 1009
Db
365 IleSerTrpTrpProAspThrIleIlePheArgSerThrAspGlyAlaThrTrpThr 384
Qy
1010 AACTGCTGGTGGCGGATGACTCATCTCCGTAGCACTCACTGTGAGCGACTGTGCTCC 1069
Db
385 ArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSer 404
Qy
1070 CCTATTGGAGTGAACAGGTATTCGAGCATCAATTACTACTACGATGATATCAGC 1129
Db
405 AlaGluProTrpLeuThrPheGlyValGlnProAsnPro---ProValProSerProLys 423
Qy
1130 AACGCCCTTGGATCCAGGACACGACTTCGACCGACCAATTCCCGGT-----CGC 1180
Db
424 LeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyr 443
Qy
1181 GTGCGTGTGATGGTGAAGCCCTTGGCATTTGACCCATTTGACTCCAACTACCTGCTCTAC 1240
Db
444 GlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGln 463
Qy
1241 GGACGGGCTGACATATATGCGGACACGACTGACCAATTGGGACTCGAAGCACAAT 1300
Db
464 IleHisIleAlaProMetValLysGlyLeuGluThrAlaValAsnAspLeuIleSer 483
Qy
1301 GTGACGCTCAAGTCACTGGCGGTGGTATTGAGGAGATGCGTGTCTCGACTGATCAGC 1360
Db
484 ProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyPheThrHisAla 503
Qy
1361 CCGCCCGGGGACCGGCTGTCTCGCTGTGTGGGACGACGCGGGGTCTTACCACCTCG 1420
Db
504 AspValThrAlaValProSerThrIlePheThrSerProValPheThrGlyThrSer 523
Qy
1421 GATCTGAGCAGCGCCCAACCAAGGCTTCCACCGCCGACCTACGCGCACCAACGCGC 1480
Db
524 ValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSer 543
Qy
1481 ATCGACTACCGCGCAACAGCCTTCCACATGTCGCGGTCCGAGTACGCGAC----- 1534
Db

544 SerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGln 563
Qy
1535 GACTATCCGAC-----CTGCTCTGTCTCGAATCTCGATCGACTTGGTATGCA 1585
Db
564 GlySerGluProGlyGlyValThrGlyThrValAlaAlaSerAlaAspGlySer 583
Qy
1586 GACTACGCGCGGTCAACAGCACCGGACCGGCGGTGGCTCTTTCCCGGATGTGAC 1645
Db
584 ArgPheValTrpAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGly 603
Qy
1646 ACGGTGCTTCTCATCTCCAGCACCTCGGGCGGTGGTCTCGAAGTCGACGGC----- 1699
Db
604 AsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgVal 623
Qy
1700 --ACCTTACTGCGGTTCCTCTCTCCGTCGCGGCGGTCTATCGCTCCGACAAAGTCC 1756
Db
624 AsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGly 643
Qy
1757 GACAACACGGTCTTCTATGTGGATCCGCGGCGCAATCTACGTGTGGAAGACACAGCG 1816
Db
644 ValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPhe 663
Qy
1817 ACCTCTTCCACAGACCGGTGAGC---CTGGGTCCAGCACCGCTCAACGCCATC--- 1870
Db
664 HisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHis 683
Qy
1871 CGCGGCATCTCTCCATCGCGCGATGCTCTGGCGGTCCACGACAAAGGCGCTCTGGCAT 1930
Db
684 SerThrAsnGlyGlySerSerTrpSerAlaIle---ThrGlyValSerSerAlaValAsn 702
Qy
1931 TCCACGGATTACGGTAGTACGTTTCAACGAGTCCGCGGCGGTCAACGCGGCGTGGAGC 1990
Db
703 ValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheValGlyThr 722
Qy
1991 TTTGGTGTGGCAAGCGCTCTCCACGGCGACTACGTCTGCTATCTACGGGTCTTTCACC 2050
Db
723 IleGlyGlyValThrGlyValTyrArgSerAspCysGlyThrTrpValLeuIle 742
Qy
2051 ATCAGCGTGGCGGGCTGTTTCAAGACGAGGACGAGGACCACTGGCAGGTGATC 2110
Db
743 AsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAla-----IleThrGlyAspHis 760
Qy
2111 TCGGATGCTGCGACCGGATTCGGCTCG---GGTCCGCCAACGTGGTGAACGCGACCTG 2167
Db
761 AlaAsnLeuArgArgVal-TyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIle 780
Qy
2168 CAGACCTACGGCGCGTGTTCGTGGGCACGAAACGCGCGGCGCATCTTCTACGGCAAT-- 2225
Db
780 eGlyGlyAlaProSerGlySerProSerProSerValSerProSerAlaSerProSerLe 800
Qy
2226 -----CCGACGGGAGCTCCCGCGGCCACGCGGCGGCGGCGGCGGCGGCGGCGG 2275
Db
800 uSerProSerProSerProSerProSerProSerProSerProSerProSerProSerPr 820
Qy
2276 ACCTCCAAGACTCG-ACGACGGTATCCAGCAGTTGAAACACCAACGCGGCGGCTC 2334
Db
820 oSerSerProSerProSerProSerProSerProSerProSerProSerProSerPr 840
Qy
2335 GACCACGCTTCCAGCACCAACCGTGAAGACCAACGACTAGCTGCTCCAGCACGACTTCAA 2394
Db
840 oSerAlaSerProSerProSerProSerProSerProSerProSerProSerProSerPr 860
Qy
2395 GGCCAGCTCGACCAACCAACAGACTACCAACCAACGAGACGACCAACGCGGCGGCGAC 2454
Db
860 oSerProThrProSerProSerProValSerGlyGly 872
Qy
2455 CACCGGACGCGCTCGCGGTAGCTCAGTGTGGCGGC 2491
Db

RESULT 10

AY040839

LOCUS

DEFINITION Aspergillus niger endoglucanase C (eglC) gene, complete cds.

3959 bp DNA linear

PLN 01-APR-2002

[illegible]

Saito, R., Sasaki, D., Sato, K., Shibata, K., Shingawa, A., Shiraki, T., Yoshino, M. and Hayashizaki, Y.

TITLE Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice

JOURNAL Science 301 (5631), 376-379 (2003)

MEDLINE 22752273

PUBMED 12869764

REFERENCE 2 (bases 1 to 2724)

AUTHORS Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Doi, K., Fujimura, T., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayashizaki, Y., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Hotta, I., Ikeda, J., Iida, Y., Ikeda, R., Imamura, K., Imotani, K., Ishibiki, J., Ishii, Y., Ishikawa, M., Itoh, M., Kagawa, T., Kanagawa, S., Katoh, H., Kawagashira, N., Kawai, J., Kawamata, M., Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M., Kodama, T., Kojima, K., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Kurosaki, T., Kusumegi, T., Li, C., Lu, M., Masuda, H., Matsubara, K., Matsuyama, T., Miura, J., Miyazaki, A., Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakamura, M., Namiki, T., Narikawa, R., Niihara, J., Nishi, K., Nomura, K., Numasaki, R., Ohneda, E., Ohno, M., Ohtsuki, K., Oka, M., Ooka, H., Otsu, N., Ota, Y., Otsu, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Satoh, K., Shibata, K., Shingawa, A., Shiraki, T., Shishiki, T., Sogabe, Y., Sugano, S., Sugiyama, A., Suzuki, K., Suzuki, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Tsunoda, Y., Ueda, M., Waki, K., Xie, Q., Yahagi, W., Yamada, H., Yamamoto, M., Yasunishi, A., Yazaki, J., Yokomizo, S. and Yoshimura, A.

Direct Submission

TITLE Submitted (27-AUG-2002) Shoshi Kikuchi, National Institute of Agricultural Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression: 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail: skikuchi@nias.affrc.go.jp, Tel: 81-29-838-7007, Fax: 81-29-838-7007)

COMMENT This clone is one of the 28K full-length cDNA clones from japonica rice.

URL : <http://cdna01.dna.affrc.go.jp/cDNA/>
NIAS Rice Full-length cDNA Project Team: Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T. and Yamamoto, M.

FAIS Genome Sequencing & Analysis Group: Otsu, Y., Iida, Y., Fujimura, T., Ikeda, R., Ishibiki, J., Kawanata, M., Kobayashi, M., Kodama, T., Kurosaki, T., Kusumegi, T., Lu, M., Masuda, H., Miura, J., Mizuno, K., Narikawa, R., Niihara, J., Oka, M., Ryu, R., Sugano, S., Sugiyama, A., Suzuki, Y., Tsunoda, Y., Ueda, M., Xie, Q., Yokomizo, S., Yoshimura, A., Matsubara, K. and Murakami, K.

Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hanagaki, T., Hiraoka, T., Hori, F., Iida, J., Imamura, K., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawai, J., Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Otsu, N., Ota, Y., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K., Yasunishi, A. and Hayashizaki, Y.

Location/Qualifiers

1. .2724

FEATURES /organism="Oryza sativa (japonica cultivar-group)"

source /mol_type="mRNA"

/cultivar="nipponbare"

/db_xref="taxon:39947"

/clone="002-167-F04"

ORIGIN

Alignment Scores:
970 ATCACTCTGTCTCCGGTAGCGACTGTACC-----TACGGATTCCGGCGCTCTCC 1020

Pred. No.: 1.88e-41 Length: 2724
Score: 1633.00 Matches: 365
Percent Similarity: 58.11% Conservative: 151
Best Local Similarity: 41.10% Mismatches: 307
Query Match: 31.80% Indels: 65
DB: 8 Gaps: 27

US-09-917-376-1 (1-957) x AK110506 (1-2724)

QY 37 ileAlaIleThrAlaSerProAlaHisAlaAlaThrThrGlnProTyrThrTrpSerAen 56

Db 88 ATCTGATACCGGCTCTCCCG-----GCCGTGTCGAGCGGCTCTACTCATCGAGAGAT 141

QY 57 ValAlaIle---GlyGlyGlyGlyPheValAspGlyIleValPheAsnGlnGlyAlaPro 75

Db 142 GCCAAGTTTCGAGCGCGGCTGTTGTCGCGGCAATTGATTCATCTCACCAGGCT 201

QY 76 GlyIleLeuTyrValArgThrAspIleGlyMetTyrArgTrpAspAlaAlaAsnGly 95

Db 202 GCGGTTCCTTATGCTCGAAGCATATAGAGCGCTGTACAGGTGAAC---AGTGATGAC 258

QY 96 ArgTrpIleProLeuLeuAspTrpVal-----GlyTrpAsnAsnTrpGlyTyr 111

Db 259 TCGTGGACGCCCATCACGACTCCATCCGACAAATGGAATGGCAATTTGG----- 312

QY 112 AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaVal 131

Db 313 ---GGAGTGGATGGCTGTCTTTGATCCCGGACCCCAATGTGTGTACTTCGCGCG 369

QY 132 GlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAspGln 151

Db 370 GGGTGTACACCACTCTGGGATCTTAGCCGCGGCATCTACAAGAGCAGCAGGACAG 429

QY 152 GlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGlyArg 171

Db 430 GGTGCGACCTGGACATCTTCAACGCTCAGCTTCAAGTCGGTGGTAACATGCTGCGCG 489

QY 172 GlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAla 191

Db 490 GGCATGGCGGCGCGCTGGCTGTGATCCCAAGAACTCCAACTTGTGTATTTCGGCGCA 549

QY 192 ProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThr 211

Db 550 CGCTCTGGAACCGGCTCTGGAAGAGCACCGCGCGGCAAGACTTCAAGAAGGTCGAC 609

QY 212 AsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGlnSer 231

Db 610 TCCTTCAAGGCTGTGTGTACCTACATCCCGTCCCGCAGCAGCAGCCGGCTTACACAGC 669

QY 232 AspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerLeuGlyGlnAla 251

Db 670 GACATCAACGGCTGAGCTTGTGACCTTGTGACCTCGCGACCTCGCAACCATCAATGGTGGC 729

QY 252 SerIysThrIlePheValGlyValAlaAsp---ProAsnAsnProValPheTrpSerArg 270

Db 730 ACTGGCGCATATTTCGTAGGCACCGCGGCAAGCTCAAGCTCGCTTTCCTGAG 789

QY 271 AspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIleProHisLys 290

Db 790 GATCCCGGAGCAGCATGGAACCGCGTCAAGGGCCAGCGCTTGGCAAAATATCTTCTCTAAG 849

QY 291 GlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyPro 310

Db 850 TGCAGGCTGCGCCACCGAAGGCGCACTGTACATCTCTTACGCGCAGGAACGGCCCT 909

QY 311 TyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGlyThrTrpArg 330

Db 910 TACGATGTGCTCTGGCGTGTCTGGCGATACCGCTCGCGGACCGGACGCGGACGGAC 969

QY 331 IleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThr 350

Db 970 ATCACTCTGTCTCCGGTAGCGACTGTACC-----TACGGATTCCGGCGCTCTCC 1020


```

/strain="QM6a"
/db_xref="taxon:51453"
/notes="anamorph: Trichoderma reesei"
1. .2517
/gene="cel74a"
CDS
1. .2517
/gene="cel74a"
/notes="putative secreted endoglucanase"
/codon_start=1
/product="Cel74a"
/protein_id="AAP57752.1"
/db_xref="GI:31747160"
/translation="MKVSRVLAIVLGVAPDAHPAFAFNKVNKVLGGGGFVPGIIFHPKT
KGVYARTDIGRLNDDSWFTVDGIADNAGNWDGIDAVLDPQDDQKYAAVG
MTNWSWDNGATIRSDRGATWSFTNLPFKVGNMPPGRGAGERLAVDPANSNIIYFG
ARSGNLGMLTIDGVTFSKVSFTAGTVPDPSNGYNSDKQGLMWTFDSTSTT
GATSRIFVGTADNITASVSVTNAGTSWAVPGQKVPHPKALQPAEKALYLYS
DGTGPDYDGLGSWRDIDAGTWDITPVSGSDLYFGFGGLDLDLQKPTLVVASLNS
WMPDAQFRSDSGTTSPIWASVYPTTYSISTPKAPWIKNNFIDVTSPPSDG
LTKRLGMIETSLIEDPTDSNHWLYGTGMTIFGGHDLTNWDTRHNVSIOSLADGIEBFS
VODLASPGSSELLAAVGGDNGFTFASRNDLGTSPQVWATPTWATSTLSDVYAGNSVK
SVRVGNTAGTQOAVLSSDGGATWSIDYAADTSMNGTWAYSDGPTILWSTASSGVQ
RSQPGSFASVSLPAGAVIASDKTNSVPYAGSGSTFYVSKDTSFTEGPKLGSAG
TRDIAAHPITAGTLIVSDVGFPRSDGTFPGQVSTALNTIYQALGVSGSNWL
YAFQTPSGARLYASDGSASMTDIQSGQFGSIDSTKVGSGSTAGQVYVGTNGRV
FYAQGTGGGTGGTSSSTKQSSSTSSASSTLRSVSVSTTRASTVTSRTSSAAGP
TSGSVAGHYAQCIGWTGPTQCVAPYVCQKQNDYYVCV"

ORIGIN
Alignment Scores:
Pred. No.: 1,94e-41 Length: 2517
Score: 1631.50 Matches: 353
Percent Similarity: 57.23% Conservative: 130
Best Local Similarity: 41.82% Mismatches: 297
Query Match: 31.77% Indels: 64
DB: 8 Gaps: 23

US-09-917-376-1 (1-957) x AY281371 (1-2517)
Qy 25 SerPheAlaValAlaAlaLeuGlyValLeuProIleAlaIleThrAlaSerProAla 44
Db 10 TCTCGAGTCTCCCTTGTCTGGGGCGGTC-----ATCCCTGGCC 51
Qy 45 HisAlaAlaThrThrGlnProTyrThrTyrSerAsnValAlaIle---GlyGlyGlyGly 63
Db 52 CATGCTGCC-----TTTTCATGAAGACGTCAAGTCGGCGGGCGGGCGGC 99
Qy 64 PheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAsp 83
Db 100 TTGCTCCCGGCATCATCTTCCATCCCAAGCAAAAGCGGTAGCATATGCACGAACAGAT 159
Qy 84 IleGlyGlyMetTyrArgTyrAspAlaAlaAsnGlyArgTyrIleProLeuLeuAspTyr 103
Db 160 ATTGGGGGGCTGTACCGCCCTCAAC--GCCGAGACTCATGAGCCCGCGTCACGGATGGG 216
Qy 104 Val-----GlyTyrAsnAsnTyrGlyTyrAsnGlyValValSerIleAlaAla 119
Db 217 ATTGCTGATAATCCCGGCTGGCAACTGG-----GGCATCGAGCGTGTTCGCTT 267
Qy 120 AspProIleAsnThrAsnLysValTyrAlaAlaValGlyMetTyrThrAsnSerTyrAsp 139
Db 268 GATCCGACGACGATCAAAAGGTGTATGCCGAGTCGGCATGTATACGAACAGCTGGAT 327
Qy 140 ProAsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTyrGlnIleThrPro 159
Db 328 CCGAGTAATGGAGCCATCATTCGCTCGTCAGACCGCGCGCAACGTGGTCTCTTCCACCAAC 387
Qy 160 LeuProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaVal 179
Db 388 TTGCCCTTCAAGTCTGGGGGTAAATCCAGGACGCGGACCGGAGAGCGTCTGGCTGTC 447
Qy 180 AspProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTyrArg 199

```

```

448 GATCCGGCCCAACTCCAACATCATCTACTTTGGTGTCTGCTCAGGAAACGGCTCTGGAA 507
200 SerThrAspSerGlyAlaThrTyrSerGlnMetThrAsnPheProAspValGlyThrTyr 219
508 TCTACGGACGGCGGTGACCTTTTCAAGGTCTCGTGGTTCCACGGCAACTGGGACGTAC 567
220 IleAlaAsnProThrAspThrGlyTyrGlnSerAspIleGlnGlyValValTyrVal 239
568 ATCCACAGCCCGAGTGATTCACACGGCTACACAGCAACAGCAAGCAAGTCTATGTGGGT 627
240 AlaPheAspLysSerSerSerLeuGlyGlnAlaSerLysThrIlePheValGlyVal 259
628 AGTTTCACCTCAACAGCAGCAGCAGCGGGGAGCCAGCTCTCGTATCTTTGTTGGCAGC 687
260 AlaAspPro---AsnAsnProValPheTyrSerArgAspGlyGlyAlaThrTyrGlnAla 278
688 GCTGATAACATCACTGCTTCACTCTATGTGAGCAGCAATGCCGGCTCCACGCTGGAGTGT 747
279 ValProGlyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsnHis 298
748 GTACCGGGGACGCCAGGGAATACTTCTCCTCACAGGCGGAAACTGCAGCCAGCAGAGAA 807
299 ValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerGlyAspVal 318
808 GCCTTGATCTGACCTATTCGATGGCAGCGGCGGTATGATGGCAGCATTGGCTCAGTG 867
319 TrrLysPheSerValThrSerGlyThrTyrThrArgIleSerProValProSerThrAsp 338
868 TGGAGGTACGACATTCGAGGGGAACTTGGAAAGACATCACTCCCTGTCTCTGGATCAGAT 927
339 ThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThr 358
928 CTA-----TACTTTGGCTTTGGCGGCTTGGCGCTCGATTTGCAAAAGCAGGAAC 978
359 IleMetValAlaThrGlnIleSerTyrThrProAspThrIleIlePheArgSerThrAsp 378
979 CTTTGTGTGCTTCTTGAACCTCTTGGGGCAGATGCTCAGCTGTTGGTGGCCGAC 1038
379 GlyGlyAlaThrTyrArgIleTyrAspTyrThrSerTyrProAsnArgSerLeuArg 398
1039 TCTGGGACAAATGGAGCCGATCTGGGGTGGGGAGCTATCCGACTCAGACCTATTAC 1098
399 TyrValLeuAspIleSerAlaGluProTyrIleThrPheGly---ValGlnProAsnPro 417
1099 TACAGCATCTCACTCCCAAGACCGCGGTGATCAAGAACAACTTTATCGATGTGACGAGC 1158
418 ProValProSer-----ProLysLeuGlyTyrMetAspGluAlaMetAlaIle 433
1159 GAGTCACCGTCCGATGGTCTCATCAAGCGCTCGGCTGGATGATTGAGTCTCTCGAGATT 1218
434 AspProPheAsnSerAspMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsn 453
1219 GACCAACCGACAGCAACCACTGGCTCTACGGCAGCGGAATGACAACTCTTTGGCGGCCAC 1278
454 AspLeuThrLysTyrAspSerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeu 473
1279 GATCTCAACAACTGGGACACGCGCACATGTGTCATCAATCACTATGTCAGACGGCATC 1338
474 GluGlnThrAlaValAsnAspLeuIleSerProProSerGlyAlaProLeuIleSerAla 493
1339 GAGGAATTCCTCGTCCAGGACCTTGCCCTCTGCACCGCGGGAAGCGAGGATATTGGCCGCA 1398
494 LeuGlyAspLeuGlyGlyPheThrHisAla-----AspValThrAlaValProSer 510
1399 GTCGGAGACGACCAACGGCTTTCACCTTTCAGCAGCAAGACCTCGGACATCGCCGCGAC 1458
511 ThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyrAlaGluLeuAsn 530
1459 ACGGTCTGGCAACGCCCCACATGGGCCACTTCGACGAGGCTCGACTACCGCGGAACCTCG 1518
531 ProSerIleLeuAlaGlySerPheAspProSerSerGlnProAsnAspArgHis 550
1519 GTCAAGAGCGTCTCGCGTGGCAACACCGCGCGGCAACAG-----1563

```


ILAVDSMESILDWYKEGMIKGGSGAGLNSLRSSKELLSGNSAGPVSFWEGA
 DASAGTIGSGGATRAAGKVLIDVDPDIEDTETIKVKEEKIRALRSDAGFMDLGGD
 DITSVOYQANNSVRVNDTFKXAVEGKFGELTSMTGVI BEVDKSLFRKWAEEAAW
 ACADPGIOYDDTINTWHHTCPESGRINSNPNCSYMHLDNTSCNLSNMLKFKDDGK
 GROSFEVERKVLWITAMDISICFADFPPTQKIGENTRAFQRLGIGYANIGALIMA
 TGHAVSDGGRALAGAITSLMTGTSYKRSAAELAAVVPYDGVARNAQPHORVMKQHSO
 ANGVAVRVDLDTPIWAATEARQDVLHLEKNGKPNRAQASVIAPGTIGLANSDDTT
 GLEPDALVKKLVGGSGMIVNGTVPOALRGLYOQEOIEAIVAHIAENGVIDAP
 GLUKHYEYFDCAMGRSISAMGHVMMAAIQWISGALSNTVNPETATVEDVEEYV
 FEAMKGVKALAIYRNCVKVQPLSAKKETEKAEVTAETATREAVEKVVYRVR
 KRLPKRPGITTSFTVGAAGYWTANSYPODGLGEVFLKMSQKSLAGMDAFSLAV
 SVGLQYGVLETVVSKFTNMRPEPAGMTDDPOVRMAQSIVDVI FRELALDLPFETRS
 ALGHSERQORHLETSPESSDDVMDVEGLAQSAPRAQELKAVATPKAEVAAVPA
 PKOATHSBELVENQIGIADAPLFCSCGTMQRAGSCYICBGGSGTSGCS"
 complement (3086. .3649)
 /note="SAV2462"
 /codon_start=1
 /transl_table=1
 /product="putative regulatory protein"
 /protein_id="BAC70173.1"
 /db_xref="GI:29606110"
 /translation="MHCPCRHPDSRVDSRTTDDGTSIRRRQCPDCSRRFTTVC
 SLWVKRSGVTPEFRTKINGVRKACQRPVTEADALQORVEAVRATGSALTT
 HVGIALGLFRLDLVAILRNASYRAFDSLEDFEAATAELEERERPAVDDEHE
 DAAERQGTDRSGGVTEVPVPAIVAD"
 4168. .4947
 /gene="lexA"
 4168. .4947
 /gene="lexA"
 /note="SAV2463"
 /codon_start=1
 /transl_table=1
 /product="putative SOS regulatory protein LexA"
 /protein_id="BAC70174.1"
 /db_xref="GI:29606111"
 /translations="MTTIDASATIQAQSRQLRLEPVHAMNEATSHGPKRSLPGRRP
 GIRADSSGLTDQRRIEYVDSVRQGYPPSREITQAVGLSSTSVAHQLMALBRK
 GFRLDPHPRAYEVRGQSQSVQPTDTAGKPAASYPLVGRIAAGGPILAESVED
 VFLPQLQVDELFLVKVVGDSMIEAAICDGMWTVRQPVVAENGDI VAAMLDGEAT
 VKRFKEDGHVLLPHNSAYQPIPGDEATILGKVVAVLRV"
 complement (5110. .7143)
 /gene="ding"
 complement (5110. .7143)
 /gene="ding"
 /note="SAV2464"
 /codon_start=1
 /transl_table=1
 /product="putative ATP-dependent helicase"
 /protein_id="BAC70175.1"
 /db_xref="GI:29606112"
 /translations="MTKPSLPPELLHAATVAVGTERPQVTVMAEVAEAIIDSSHLIV
 OAGTGTGSLVLPALVAPALHGERVIVATATLALORQLVERDLPRTVDALHPLRRPEF
 AMLKRSNYLCILHRLHEGVQDREELGFPQFEAAAPTSLKQDLRLRWSDETETGD
 RDLTTPGSDRAWAQSVSSRECLGATKCAYGACFAESRERAKLAEVVVTHALLA
 IDAEGAPVLPOHEVILVDEAHLEISVGTGATGELTPOGVNRAAKLVNNEKAD
 QLOTAAGFERLMELALPGRLEIPEDLGALMALDAAKTIVSAICATKDKSQVED
 AVRKOALASVESVGDVNRVWYERHDFRFGASLRVAPMSVGLLRKLFPA
 DRSVLTSLTKLGGPVGWASIGLAPETQGDLLPQMKGVVGSFPFYPKGILYV
 AHLSPARDGDRADMDELITELIQAGRTILGFSMRAQALAAELSRIFEFPL
 LQGEETLGLKNFADPTCLFGLTSLMQGVDPVPGSQLVVMKIPFPDPDPLMS
 ARQKAVEDAGGNGFMAVATTHALLMAQAGRLVRASGDRGVAVVADLORLATYGSY
 LKASLPDFWYTTDRNOVRRSLTAIDEAAKTEVTEVTEVAEVEEAAKESGA"
 complement (7376. .8509)
 /note="SAV2465"
 /codon_start=1
 /transl_table=1
 /product="hypothetical protein"
 /protein_id="BAC70176.1"
 /db_xref="GI:29606113"
 /translations="MPPTDASTDAEPATSDPARIASARSAAEDSTLELRPLDELV
 ALFPAEEAAGGEAAGGVAGVREGTAGGEEDAVGGAEFPAGRAGDAGEAQTATA
 APSVPAARATASADSGSDAVDPAAPFAASTVPSASRVPSPSVTSAPSVTAT
 APAPAPPSLGDTKALAGQAVSVTRAADDLLGRVEDWGPTVTPAGTFLQAPVPERDL

PLLRWNNDPAVAFWKLADGESVTEQHLRAOLGCDGSRSPVCLGVLECTPMSYWEIYR
 ADDLSARHPYPARPHDTGHIHLIGVADRGRGLSTLLRAVADVLDRRPPSCARVAE
 PDLRNTSSVSFAFLGAGFRFAEVDLPDKRAALMVRDRPLROLL"
 complement (8549. .10609)
 /note="SAV2466"
 /codon_start=1
 /transl_table=1
 /product="hypothetical protein"
 /protein_id="BAC70177.1"
 /db_xref="GI:29606114"
 /translations="MNAATPTSDGSRTHERRAGHDTQPTVOGPOQVAFASEAQAAPAE
 TOAPLAPATOVVFVEVTRTPVQASESPVPVGRARSERLRGTTADPLDHPAOTAA
 QAAVENLLSCWRENGLVAPDDGTLRIPLLATGTALLVPVHHSPTGWHREGLPYLA
 AVFIDRRRAPADPEPDLFAEQALLLHPLHPTFKREGISEAEARLYSPELRGSFA
 LHMVAAPVLATNSAINTERGRPVPAEQTLARLAGDGLPMDPGDFAALPLHPWQIREIR
 HRPEAAALDAGLIRDLGHPGPAWHPTSSVTRHTGAPAMLKLSGLRIITNSREIR
 RKLHGVVEVHRLRSGLGEQOAHQPFIDVRDPAVLAVDDQHGNAVGLDVMIRHN
 PFAPTDVSCVAGLVSPRPWFOEQRPVWRSRLAEIIRLAGRTGRPGAVATEWF
 LRYLEQVVRVPLMDSSEAGLAEHQNTLLLDPEGMPTGGRYNDNQGYFRTSRA
 ELERLPGIGEHSDTFVSDEVTRFAYLGINNVGLIGAFGSHLAEQQLLAAR
 RFLADVATGPARLTPLTRLLDSPVLRCRKNALLTLHLGLDELVGPVDTQSVYVTIAN
 PLHS"
 complement (10775. .12229)
 /note="SAV2467"
 /codon_start=1
 /transl_table=1
 /product="putative class III aminotransferase"
 /protein_id="BAC70178.1"
 /db_xref="GI:29606115"
 /translations="MAVTGASVASKREAGAHAEHILRRQSRARESAARTYARALP
 IVPFRAGLTIEGAGRRYLDLCSGAGTALGHNHVPVLEAIRKVLDSGAPLHVLDLA
 TVKDAFTTELFTLPAGLADRARVQFCGPTGDAVEAFKLVRTATGRSGLIAPFTA
 YHGTAGALEASGATDGRVARLPVPOYPCFPFGIGGERGAELGARWTESLDDVKSQ
 VPLPAGMILEPVOGEGGVPAPDAWLRMRAITADRSIPLIADSVQTVGRTGTFWAV
 EHSVVDPAWVLSKAIGSLPLAVVYVDDLVQVQPHAGTFRGNQLAMAGATATLA
 YVREHRAAQITIGARMLDQLSLATEFPCVGEVRGMLGIMVSELVRAEEDREAVSP
 GSGDFASLDTAGAGPRPAAPAEAAVQRECRRLGILVELGGRHASVVRLLPLTISDE
 QATVADRLADAVEAVAGRAERSRGFORSERAG"
 12761. .13993
 /note="SAV2468"
 /codon_start=1
 /transl_table=1
 /product="putative secreted protein"
 /protein_id="BAC70179.1"
 /db_xref="GI:29606116"
 /translations="MRPIRLFAARGESABRRTSPVLGAVGLATATALTATACDSGD
 DTKASSEFSASAGDDKIPDIRDLKXRGIDIDKWNKNGAKWNRRDMLREAN

Alignment Scores:

Pred. No.:	1.29e-37	Length:	299550
Score:	1570.00	Matches:	393
Percent Similarity:	49.25%	Conservative:	131
Best Local Similarity:	36.94%	Mismatches:	346
Query Match:	30.57%	Indels:	194
DB:	1	Gaps:	38

US-09-917-376-1 (1-957) x AP005031 (1-299550)

Qy	4	SerGluAsnIleArgLeuThrMetArg	-----SerArgArgLeuValSerLeu 19
Db	161525	TCTCCAGAACGAGGATGTATGTCGCGACGCCGCCGCCGCGACGACCAACCGTC	-----161472
Qy	20	LeuAlaIleThrAlaSerPheAlaValAlaAlaLeuGlyValLeuProIleAlaIle 39	
Db	161471	CTCCCGGAGCCCGCCGCG	-----GCCGCGCGCTCAGCGCGTCCCC-----GCC 161427
Qy	40	ThrAlaSerProAlaHisAlaAlaThrThr	-----GlnProTyThrTrpSerAsn 56
Db	161426	GTGGGGTCCAGCGCGACGCCGCCGAAACCAACCGCGCGCGCTCGTACCGCTGGCGCAC	161367
Qy	57	ValAlaIleGlyGlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGly 76	


```
Qy 770 ThrAsnGlyArgGlyIleValTyrGlyAspIle-----GlyGlyAlaPro 784
Db 159323 ACCAACGGCGCGGCATCAGTACGGGAGCGCGTCTGTGATGCCGGGCTCACGGATGCCA 159264
Qy 785 SerGlySerProSerProSerValSerProSerAlaSerProSerLeuSerProSerPro 804
Db 159263 CCGGGGCGCGATCTCTGTTGGTGGCG---ACTACAAACCCGAGACAGTGGCCCGAGGAGG 159207
Qy 805 Ser------ProSerSerSerProSerPro----- 812
Db 159206 TCTGGCAGGAGGACGTACGCCTCATGAGGAGCGCGCGTCAACTCGTCAACCGTCGGCG 159147
Qy 813 ----SerProSerProSerSerPro----- 820
Db 159146 TCTTCTCTCTGGCGGAGCTCGAACCACCCCGGGCGGAGGAGTTCGGTGGCTCGACCC 159087
Qy 821 ----SerSerProSerProSerPro----- 828
Db 159086 GGTCTATGACCTGATGCAGAGAACGGCATCGGCGTCTGCTCGGCCACCCACCGCCT 159027
Qy 829 SerProSerPro----- 832
Db 159026 CGCCCCCGCCTGGATGGGGCGCTGCACCCCGAGACCTGCCCGCGAGGAGCGGCC 158967
Qy 833 ----SerProSerArgSerProSerProSerAla 842
Db 158966 GCATCGAGTGGTGGCGCGCGCCAGCACCTCGCCCACTCCAGTGGCGCTTACCGGCGCT 158907
Qy 843 SerProSerProSerSerProSerProSerSerSer-----ProSerSer 858
Db 158906 ACCCGCGCGCATCACCGAGACCTCGCCCGCGCTACGGCGGCGCATCCCGCCCTCACGA 158847
Qy 859 SerProSerProThrProSerSerProValSerGlyValIle----- 874
Db 158846 TGTGGCACATCACACGAGTACTGTGACCTACGACTGGGCGGAGCGCGCCACCGCCT 158787
Qy 875 ----ValGlnTyrIleAsnAsnAspSerAlaProGlyAsp 886
Db 158786 TCCGCGCGTGGCTCCAGGCGGAGGTACGGCAGCTCGACGCTCAACACGCGCTGGG--- 158730
Qy 887 AsnGlnIleIlePro---GlyLeuGlnValValAsnThrGlySerSerSerValAspLeu 905
Db 158729 ----GCACGGCGCTTCTGGAGCGAGGCTACGACGACTGGGAGGCGATCTCCCGCGCG 158676
Qy 906 SerThrValThrValArgTyrTyrPheThrArgAspGlyGlySerSerThrLeuValTyr 925
Db 158675 GCCACGCGCCTACATGAAACACCCACCGAGGTGCTGACTTCAAGCGCTTACCTCCG 158616
Qy 926 AsnCysAspTrpAlaAlaIleGlyCysGlyAsnIleArgAlaSerPheGlySerValAsn 945
Db 158615 ACGCACTCATGGAGTGTCT---TCGTGCGCGAAGCTGACATGCTG----- 158574
Qy 946 ProAlaThrPro 949
Db 158573 CCGGCGACACCC 158562
RESULT 14
LOCUS AB011809 14520 bp DNA linear BCT 29-MAY-2002
DEFINITION Xanthomonas axonopodis pv. citri str. 306, section 187 of 469 of
the complete genome.
VERSION AB011809 AB008923
KEYWORDS AB011809.1 GI:21107967
SOURCE Xanthomonas axonopodis pv. citri str. 306
ORGANISM Xanthomonas axonopodis pv. citri str. 306
Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
Xanthomonadaceae; Xanthomonas.
REFERENCE 1 (bases 1 to 14520)
da Silva, A.C.R., Ferro, J.A., Reinach, F.C., Farah, C.S., Furlan, L.R.,
Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida
Jr., N.F., Alves, L.M.C., do Amaral, A.M., Bertolini, M.C.,
Camargo, L.E.A., Camarotte, G., Cannavan, F., Cardozo, J.,
Chambergo, F., Ciapina, L.P., Ciccarelli, R.M.B., Coutinho, L.L.,
Cursino-Santos, J.R., El-Dorry, H., Faria, J.B., Ferreira, A.J.S.,
Ferreira, R.C.C., Ferro, M.I.T., Formighieri, E.F., Franco, M.C.,
Griggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite
Jr., R.P., Lemos, E.G.M., Lemos, M.V.F., Locali, E.C., Machado, M.A.,
Madeira, A.M.B.N., Martinez-Rossi, N.M., Martins, E.C., Meidanis, J.,
Menck, C.F.M., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T.M.,
Okura, V.K., Oliveira, M.C., Oliveira, V.R., Pereira Jr., H.A.,
Rosi, A., Sena, J.A.D., Silva, C., de Souza, R.F., Spinola, L.A.F.,
Taki, M.A., Tamura, R.E., Teixeira, E.C., Tezza, R.I.D., Trindade dos
Santos, M., Truffi, D., Tsai, S.M., White, F.F., Setubal, J.C. and
Kitajima, J.P.
Comparison of the genomes of two Xanthomonas pathogens with
differing host specificities
Nature 417 (6887), 459-463 (2002)
2022145
12024217
2 (bases 1 to 14520)
da Silva, A.C.R., Ferro, J.A., Reinach, F.C., Farah, C.S., Furlan, L.R.,
Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida
Jr., N.F., Alves, L.M.C., do Amaral, A.M., Bertolini, M.C.,
Camargo, L.E.A., Camarotte, G., Cannavan, F., Cardozo, J.,
Chambergo, F., Ciapina, L.P., Ciccarelli, R.M.B., Coutinho, L.L.,
Cursino-Santos, J.R., El-Dorry, H., Faria, J.B., Ferreira, A.J.S.,
Ferreira, R.C.C., Ferro, M.I.T., Formighieri, E.F., Franco, M.C.,
Griggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite
Jr., R.P., Lemos, E.G.M., Lemos, M.V.F., Locali, E.C., Machado, M.A.,
Madeira, A.M.B.N., Martinez-Rossi, N.M., Martins, E.C., Meidanis, J.,
Menck, C.F.M., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T.M.,
Okura, V.K., Oliveira, M.C., Oliveira, V.R., Pereira Jr., H.A.,
Rosi, A., Sena, J.A.D., Silva, C., de Souza, R.F., Spinola, L.A.F.,
Taki, M.A., Tamura, R.E., Teixeira, E.C., Tezza, R.I.D., Trindade dos
Santos, M., Truffi, D., Tsai, S.M., White, F.F., Setubal, J.C. and
Kitajima, J.P.
Direct Submission
Submitted (28-Nov-2001) Departamento de Bioquimica, Universidade de
Sao Paulo, Av. Prof. Lineu Prestes 748, Sao Paulo, SP 05508-900,
Brazil
FEATURES
source
1. 14520
/organism="Xanthomonas axonopodis pv. citri str. 306"
/mol_type="genomic DNA"
/strain="306"
/db_xref="taxon:190486"
/note="pathovar: citri"
87. 2342
/gene="celA"
/note="synonym: XAC1770"
87. 2342
/gene="celA"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/trans_table=11
/product="cellulase"
/protein_id="AA036634.1"
/db_xref="GI:21107968"
/translaton="MIAVIDPPTKRFSAHAWLTAVCLTLLFWTSEASDEPCTP
DPYQWHSVAGCGGFTVGLFHPAERDLAYARTDVGAYRWDAQOQWALDMLGAD
DNLNGIDAFADPADPNALYLAAGTMYHERAGNAVLRSFDRGKQFGRADLPKLG
NQLGRANGERLAVDPHGRVLLGSRDGLRSDRGARHARVEVGFPPADALAGATRN
HVGSEQVAGIAPVVFDAASGHAGATPRIYGVSTAQTSLYVSEDAGRTWSAVGQPK
GLRPSHMVGHNAGWYLSYGRDPGLMAGALWNYDATQGRWRISIPQPATGDFG
GWGAVADPQHPVLLASTFRRTPRDEVRSGDGRSWVPLLAQDPHSAAPWTAH
ATPHWIGALADFPDSNHTFTVGTIWSRNLTQTFSEQQFLQWFOQRLGETVPLD
LLSPMAGALLSALGIDGFRHLDLDTAQLQYLPRLTNGESIDGAGQAPQWVRSGT
VDRNRNEIRALYSQDGGTHWAFSEPRGCGAGTIAIADASQVWVVDGQGWRT
GDFGRWQVQLPDTAVVADRVDAQRWAAADRVSGLYSDSGASPRDGTQGVGS
PADERARPOLRPPDPRWAGVYVLAAPTIGVMRWONGOLRTLSKSEARSLGKALRA
GAPPALYLAGRVAGVYDGIERSDDEGAHWRINDDAHRFGKPYSTGDPRIAGRVFAT
GGRGIFYGDP"
/gene="XAC1771"
```



```
Qy 147 ArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGly 166
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 507 CGTTCTGTCGATCGCGTAAGCAGATTGCAACGGCGCGACCTGCGCGTTCAAGCTCGGTGGC 566

Qy 167 AsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIle 186
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 567 AATCAATTGGCGCGCGCCCAATGGCGAGCGCTGCGCGTTGACCGCACGACGCGCGGTG 626

Qy 187 LeuTyrPheGlyAlaProSerGlyLysGlyLeuTyrArgSerThrAspSerGlyAlaThr 206
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 627 CTGTGCTTGCTTCACGTGATGCG---GGCTGTGGCGGACGACGATCGCGGGCGCAC 683

Qy 207 TrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThr 226
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 684 TGGCGCGCGGTGGAAGGATTTCCTCGCGATGTCATTGGCTGGCGGCCACGCGCGCAACCAT 743

Qy 227 ThrGlyTyrGlnSerAspIleGlnGlyValIleValTrpValAlaPheAspLysSerSer 246
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 744 GTCGGAAGCAACAGCGCGTG---GGCATCGCTTCGTGTGTTCGACGCGCGCCAGCGGC 800

Qy 247 SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProVal 266
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 801 CATGCTGGCCCTCGACACCGCATATGTGGCGCTTTCGACTGGCGCAACCCAGCGCTG 860

Qy 267 PheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPhe 286
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 861 TATGTCTCCGAAGATCGCGGCGCACCTGTGTGGCGGTGGCGGCAACCCAGCGGCGTG 920

Qy 287 IleProHisLysGlyValPheAspProValAsnHisValLeu----- 300
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 921 CGCCCC-----AGCCACATGTCGTGCACAAACGTCGGGCAG 956

Qy 301 ---TyrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspValTrp 319
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 957 TGTACTTGTAGCTATGGCGATGCGCCCGCGCGATCTGATGGCGCGTGGCGCGCTGTGG 1016

Qy 320 LysPheSerValThrSerGlyThrTrpThrArgLysSerProValProSerThrAspThr 339
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1017 AACTACGACGCGACGAGCGCGTGGCGGAGATCAGCCCGATCCCAACACGACCCACT 1076

Qy 340 AlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIle 359
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1077 ---GGCGATGTTTCGCTGCGCGCGCGTGGCGGTGGATCGCAACATCCCGCGAGTGCTG 1133

Qy 360 MetValAlaThrGlnIleSerTrpTrpProAspThrIleIlePheArgSerThrAspGly 379
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1134 CTGGCCAGCACGTTCCGCCCGCGCACCGCGCGGATGAGGTGTTTCGACGCGCGATGGC 1193

Qy 380 GlyAlaThrTrpThrArgIleTrpAspThrTrpThrSerTyrProAsnArgSerLeuArgTyr 399
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1194 GGACGAGCTGGGTTCGCTGTTGGCGGTGGCGGATTC----- 1232

Qy 400 ValLeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnProProVal 419
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1233 ---GATCACAGCGCGCGCGGTGG---ACCGCGCACGCAACACCGCAC----- 1274

Qy 420 ProSerProLysLeuGlyTyrMetAspGluAlaMetAlaIleAspProPheAsnSerAsp 439
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1275 ---TGATAGGC---CGCTGGCGGATCGATCGCTTCGATAGCAAC 1313

Qy 440 ArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAsp 459
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1314 CAGCCACGCTGTCACCGGTACCGGATCTGGGCTTCGGCGATCTGCAGACGCTTTCA 1373

Qy 460 SerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluGluThrAlaValAsn 479
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1374 GAGCAGCAGCGCTGCAGTGTGGTTCACGAGCCGTTGGGTGGGAAGAAACGTCGCCGCTG 1433

Qy 480 AspLeuIleSerProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGly 499
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1434 GATCTGTCAGCCGATGGCGCGCGCATCTGTCTACGCGCTGGCGGATATCGATGGC 1493
```



```
gene
4555. .6747
/genes="c1rA"
/notes="synonym: XCC1750"
CDS
4555. .6747
/genes="c1rA"
/notes="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/genemark"
/codon_start=1
/transl_table=11
/product="TonB-dependent receptor"
/protein_id="AAM41041.1"
/db_xref="GI:21112843"
/translation="MUPPDLGKPCWDYHASTSPRLPRPDVNAATTGGQVRAPARQK
IESTCVLGSTWSVQHRHHPAGVAGQSRIFRFRSVMASIVATLLSAQYADATTT
APASDATQQLDTQVHTTRSGVTKAQLVQNAEIVDSIVAEDIGKLPDNNVAEAL
ORISQIGSNRYGEGSSIAIRGLTVQRTLNGRDIETANDGRGLSPEDVSAELGSGV
VYKNSADMLEGSLGQVTLRLPELPEYGRKIAGSVQVNYVDLADGKPAFGLFSN
RWQTEIGETGILVNYQQSQSPFRQDTISIEPWTQTDLPYEGCQGSVPHGAGINTI
GRERKIGAFAMQWRNDSVEVYTVQLRSDYFSQWDSYFFALTQRNPGLPGIQVN
DRNEFYGFQNFYTESNTSLTERHSVTTDYSLGAKWTNEKLTLSLTDYQYVDATTG
TRYIVATGQDTSQFNVDFRGLPRLSDASGEGYLDVNNYDGRVHLDNDKNK
GTEFAWADMADLAFDSDFVRSPKAGVRYTDRDAETKGNTRYFMCLNTACGAATDSPA
APPSIGLVNPIPIIDFRGSSSTFGRITLTAANSVSNYEQTATFGASPLVFAENLNA
QTEKTYAAGVLRFVDSISDPFDGNIGVVRVTEVSGNVRGTDAEGGGLIPVAHQ
QTYTDVPLSLNLRMLSDQLQWRFPAASRGISRPTFDRLNPNLSLSTGT"
6853. .7455
/genes="XCC1751"
/notes="XCC1751"
CDS
6853. .7455
/genes="XCC1751"
/notes="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/genemark"
/codon_start=1
/transl_table=11
/product="TonB-dependent receptor"
/protein_id="AAM41042.1"
/db_xref="GI:21112844"
/translation="MMYGLTFYKKVEGFIANVFNVEYDQVWQITRPVNGDAGKING
AELGYTFDFGLFSGFGLQNTYTVDSSEAPSPATDNLGQSLTVPLEGLSKNSYN
AILSYETARFQGRVAVNWRSDMLVTTAGNCTGNLVPVYKNGFGQLDASLRNINDVWSI
SLDGNLLDTRRSYLGTSRYRDFVINDRRYGLTLASL"
7567. .9822
/genes="XCC1752"
/notes="XCC1752"
CDS
7567. .9822
/genes="XCC1752"
/notes="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/genemark"
/codon_start=1
/transl_table=11
/product="cellulase"
/protein_id="AAM41043.1"
/db_xref="GI:21112845"
/translation="MIAMIDPPFKRRTTTTGAWLAVLSLLLLLFTSPNVAEPATSG
PQWRSVAGGGGFTVGLVLFHPAERGLAYARTDVGGAYRWDAQOQTALTDWIGADD
WNLMDGDAVADPADADALYLAAGTYMHERAGNAVLRSFNRGRTFERADLPFKLGN
QLRANGERLAVDPHGRVLLGLSRDAGLWRSDDRGHAKVASFPDAALAGATARNH
VQREOAVGTAFVYDASNGTPTPRIYVGVSTEOTSLVKSVDAGRSWAPVAGQPRG
LRPSHWAGSDGHWLVSQDQPCGDLMACGALKWFTPAQGRWEI SPIQOPASGDGFG
WGAVADPQOPQVLLASTFRRTPRDELRSVDCGKHAPLLADAVFDHSAAPWTAHA
THWMGALAIDDPDGNHAFVTVGYIWSRNLDQFAAPLPQWFOFDRGLEETPLID
LPSMAGALLSALGIDIGDFHDELDRALQYAGPRLTNGESIDAAGQAPQWYVRSOT
VDRNRNEIRALYSRGGQKWTAFAPSEPPAGQAGSIATGADAQVWVWAPERGWNRT
SDPAQWQVVDGIPDPAVMADRVADRVYADVAGSGLYESTDAARSPRATGVGVGS
PARDETRQLQRPDPHRAQVYVLASPGKGVWRWQDGTQLVLSQPDARSLSIGKALRA
GAPPALYAGRVGVGVDFRSDGGVQWQKINDDAHRFRGPRSVTGDPRTAGRVFAT
GGRGIFYGDP"
9926. .11839
/genes="XCC1753"
/notes="XCC1753"
CDS
9926. .11839
/genes="XCC1753"
/notes="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/genemark"
/codon_start=1
/transl_table=11
/product="sialic acid-specific 9-O-acetyltransferase"
/protein_id="AAM41044.1"
/db_xref="GI:21112846"
/translation="MTPSHISIRACLLGLVLLSSGSAQAQPSLPLFADGAVLQDRDAMP
VWQWATPGASITVSDGCKTAKTDAQOWKVLPAHTAGGPPYVLSVHGDDGALQVRD
VLVDVWLASGQSNMWPQAQSDGPOAARAANDQLRHFHKVPKWSVQPETKLTGGD
WKAAMPENAGAFSAVFFPAKELRASPVGPIVDSTWGGSALEAMDAASLGLKAEQ
NQGAIEAKORDAAQAATKRIGRWPKVDGTPQWRTPEDFDDWDWDSIPVTQQWELS
GYEGMDGIWYRTTFTLTAABAKAGVQGVQIDSDITYYNGQVQGVQTEKQWNLPRV
YAVPAALRAGVQIAVRVEDLSGGGGMHGDQRFVQIGSTQALACGKFRPAQVRV
SLTDNKNOLPTLVNOMIHPLOPPVKGVIWYQGETNATDCAVYKQDQFAMIQOVR
AERGAELPLWQLANFKAGGDTGDLSPWALLRESQSKTLALPATGQALIDIGNPT
DIPTNKRQDVHRLALAAAHVAYGETLVYSPAFARASYADQASVQDQLQGSALQVR
EGGTLQGFAGVAGADKRFHPAQARIGRDIRVIRSDAFAVAVRVYGSNPADANLVNR
EQLFVSPFRTDTW"
ORIGIN
Alignment Scores: 7.12e-27 Length: 11910
Pred. No.: 1170.50 Matches: 299
Score: 1170.50
Percent Similarity: 48.61% Conservative: 138
Best Local Similarity: 33.26% Mismatches: 350
Query Match: 22.79% Indels: 113
DB: 1 Gaps: 28
US-09-917-376-1 (1-957) x AB012276 (1-11910)
QY 23 ThrAlaSerPheAlaValAlaAlaLeuGlyValLeuProIleAlaIleThrAlaSer 42
Db 7606 ACCAACAAGAGAGATCGCTGCTGTGTAGACCTGCTG---CTGTTGCTTTTCGACG 7662
QY 43 Pro-----AlaHisAlaAlaThrGlnProTyrThrTrpSerAsnValAla 58
Db 7663 CCCAACGTGCTGCGCGCGAGCCCGCCGCGCTACCACTGGCGCAGTGTGCG 7722
QY 59 IleGlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeu 78
Db 7723 ATTGGCGGTGCGGATTTGTACCGGTGTGCTGTTTTCATCCGCGCGAACGTGCTGCGCC 7782
QY 79 TyrValArgThrAspIleGlyGlyMetTyrArgTTPAspAlaAlaAsnGlyArgTrpIle 98
Db 7783 TATCGCGCACCGATGTGGTGGCGCGTACCGCTGGATGCGCGCAGCGCAGTGGACC 7842
QY 99 ProLeuLeuAspTrpValGlyTyrAsnAsnTrpGlyTyrAsnGlyValValSerIleAla 118
Db 7843 GCGCTGACCGACTGTGTTGGGCGCGCGACGACTGGAACCTGATGCGCATCGACCAITCGCC 7902
QY 119 AlaAspProIleAsnThrAsnLysValTrpAlaValGlyMetTyrThrAsnSerTrp 138
Db 7903 GTCGACCCCGCGATGCGCGTGTATCTGCGCGCGCGCACCTATATGACCAACGC 7962
QY 139 AspProAsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThr 158
Db 7963 GCGCGCAAT---GCGCGAGTCTGCTGCTGTTCAACCGCGCGCGCACGTTCCGCGCGCC 8019
QY 159 ProLeuProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyClnuArgLeuAla 178
Db 8020 GACCTGCGTTCAACTGGTGGTGAACACGCTGGCGCGCGCAATGGCGAGCGGTGGCG 8079
QY 179 ValAspProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTrp 198
Db 8080 GTGACCCACACGATGCGCGCGTGTGCTGCTGCGCGCGCGATGCC---GCGCTGTGG 8136
QY 199 ArgSerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThr 218
Db 8137 CGTAGCGACGATCGCGCGCGCACTGGCGGAAAGTGGCTGCTTTCCGACCGCGCGTG 8196
QY 219 TyrIleAlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTrp 238
Db 8197 GCGCGTGCCACCGCAGCAATCATGTTGGCGCGCGAGCGCGGTG---GGGATCGCTTT 8253
QY 239 ValAlaPheAspLysSerSerSerLeuGlyGlnAlaSerLysThrPheValGly 258
```

Db 8254 GTGCTGTTGACGACCGACAGTGGCAACACCGGCACCGCCAAACACCGGCATCTACGTGGGC 8313
Qy 259 ValAlaAspProAsnAsnProValPheTrpSerArgAspGlyAlaThrTrpGlnAla 278
Db 8314 GTGTCCACCAACAGACAGCCAGCTGTATGTGTCGAAGATCCGCGCGCAGTTGGCCACCG 8373
Qy 279 ValProGlyAlaProThrGlyPheIleProHisGlyValPheAspProValAsnHis 298
Db 8374 GTGCGCGGCGACGCGCGCGCTGCGCCGAGGACACATGGCCGCGCGACGATGGGCAC 8433
Qy 299 ValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspVal 318
Db 8434 ---TGTACTGAGCTACGGCGACACCGCGCGCGACCTGTATGGCGGGGAGCGCTTG 8490
Qy 319 TrpLysPheSerValThrSerGlyThrTrpThrArgIleSerProValProSerThrAsp 338
Db 8491 TGGAAATTCACGCGCGACAGGGGCGCTGGCGGAGATTAGCCCGATTCGCGAG---CCA 8547
Qy 339 ThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThr 358
Db 8548 GCCAGTGGCGATGGATTGCGTGGCGCGCGTGGCGGTGATCCGCAACACCGCAGGTG 8607
Qy 359 IleMetValAlaThrGlnIleSerTrpTrpProAspThrIleIlePheArgSerThrAsp 378
Db 8608 CTGTGGCCAGCACCTTCGCGCGCGCACCGCGCGACGAGCTGTATCGACGCTGGAT 8667
Qy 379 GlyGlyAlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArg 398
Db 8668 GGTGGCAAGCACTGGCGCGCGTGTG---TTGGCGCGAT 8700
Qy 399 TyrValLeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnProPro 418
Db 8701 GCGGTGTTCGACACAGCGCGCGCGTGG---ACCGCACATCCACCGCGCAC----- 8751
Qy 419 ValProSerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspPropheAsnSer 438
Db 8752 -----TGGATGGGG---GCGCTGGCGATCGATCCGTTCGACGCGC 8787
Qy 439 AspArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrp 458
Db 8788 AACACGCGCTGTTCGTACCGGCTACGGCATCTGGGCGCTCGCGCAATCTG----- 8838
Qy 459 AspSerGlyGlyGlnIleHisIleAlaPro-----MetVal 470
Db 8839 -----CAGATTTCGCGCGCACCGACGCGCGCTGCGAGTGGTGTTCAGGAC 8886
Qy 471 LysGlyLeuGluThrAlaValAsnAspLeuIleSerProSerGlyAlaProLeu 490
Db 8887 CGTGGGCTGGAAGAAACCGTGGCGCTGACCTGCTCAGCCCGATGGCGCGCGCATCTG 8946
Qy 491 IleSerAlaLeuGlyAspLeuGlyPheThrHisAlaAspValThrAlaValProSer 510
Db 8947 CTCAGGCGCATCGCGCATATCGATGCTTCGCGCATGACAGATTGGACCGCGC---CAG 9003
Qy 511 ThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyrAlaGluLeuAsn 530
Db 9004 TTCAGTACCGCGCCACCGCTGACCAATGGCGAAGCATCTGTCGCGCGCCAGGCG 9063
Qy 531 ProSerIleValArgAlaGlySerPheAspProSerSerGlnProAsnAspArgHis 550
Db 9064 CCGCAGTGGGTGTGGCGAGCGGTACCGTG---CGCGACCGCGCAACAGCAATCCCGC 9120
Qy 551 ValAlaPheSerThrAspGlyGlyLysAsnTrpPheCln---GlySerGluProGlyGly 569
Db 9121 GCGCTGATTTCGCGTATCGCGCAAGCAATGGCGAAGCATCTGTCGCGCGCCAGGCG 9180
Qy 570 ValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaPro 589
Db 9181 GGGCAGGCGCGCGCAGCATGTCATTTGGTGGCGATGCGCGCGAGTGGTGGTGGCACC 9240
Qy 590 GlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSer 609
Db 9241 GAGCGCGCGCGCAAT-----TGGCGCACGCTCCGACTTCGCGCGCAGTGGCAGCGTGTG 9294

Qy 610 GlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyr 629
Db 9295 GACGCGCTGCGCGACACCGCGGTGTGTATGTCGCGATCGTGCAGCGACGCGCTGTAT 9354
Qy 630 Ala-----LeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGln 647
Db 9355 GCGTGATGTGCGCGCGCGCATGTACGAGACCGATGCGCGCGCGCATTTCCGT 9414
Qy 648 ProValAlaAlaGlyLeu----- 653
Db 9415 -----GGACCGGTGTGACGTGGCGACCGCGCGCGCGACGCGCACCGCGCGCAA 9468
Qy 654 -----ProSerSerGlyAlaValGlyValMetPheHisAlaValProGlyLysGluGly 671
Db 9469 CTGCGTCCGACCGCGTGGCGCGAGCGTGTGTATCTGGCCAGCCCGCGCAAG----- 9522
Qy 672 AspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSerThrAsnGlyGlySerSerTrp 691
Db 9523 -----GGTGTGATGCGCTGGCAGGACGGT-----ACGCTG 9552
Qy 692 SerAlaIleThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGly 711
Db 9553 CAGGTGCTCTCACAGCGCGACGAGCGCGCTGCTCGGCGATCGGCAAGCGCTCGCGGCC 9612
Qy 712 SerSerTyrProAlaValPheValValGlyThrIleGlyGlyValThrGlyAlaTyrArg 731
Db 9613 GGTGCGCGCGCGCGCTGTATCTGCGCGCGCGTGCAGGGGTGGATGCGCTGTTCCGC 9672
Qy 732 SerAspAspCysGlyThrThrTrpValLeuIleAsnAspGlnHisGlnTyrGlyAsn 751
Db 9673 TCCGACGATGGCGCGCTGACGTGGCAACGATCAACGATGACGCGCACCGTTCGGGCGC 9732
Qy 752 TrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgValTyrIleGlyThrAsn 771
Db 9733 ---CCGTACAGCGTACCGGTGATCCCGCGCATTCGCGCGCTGTGTACTTCGCCACCGC 9789
Qy 772 GlyArgGlyIleValTyrGlyAspIleGlyAlaProSerGlySerProSerProSer 791
Db 9790 GCGCGCGCATTTCTACGCGCATTCGAGATGCCGATGTCATGCTGTCTGTCGCGA----- 9843
Qy 792 ValSerProSerAlaSerProSerLeuSerProSerProSerProSerProSerPro 810
Db 9844 ---CGCCTGTGCTTTTACCGTGTACGCGCGCGCGCGCTGCGCGCGCGCTGCA 9900
Qy 811 -----SerProSerProSerProSerProSerProSerProSerProSerPro 827
Db 9901 CACGCGCGACCTCGAGATCACCGCATGACCCCGT---CCACATTTTCCGCGCTGTTCG 9959
Qy 828 -----ProSerProSerPro----- 832
Db 9960 TGGGTGTGTGCTGTCTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 10019
Qy 833 -----SerProSerArgSerProSerProSerProSerProSerProSerPro 846
Db 10020 CCGATGGCGCGGTCTGACGCGCGATGAGGCGATGCCGATGGGCGTGGGCGCGCGCG 10079
Qy 847 SerSerSerProSerProSerProSerProSerProSerProSerProSerProSer 865
Db 10080 GCGCTTCAA---TCACGCTGAGTTTCGATGTGCAAGCAGCGCGCACCGCGCGCGCG 10133

Search completed: May 14, 2004, 14:54:22
Job time : 9874 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 14, 2004, 08:53:04 ; Search time 848 Seconds
(without alignments)
4794.252 Million cell updates/sec

Title: US-09-917-376-1
Perfect score: 5135
Sequence: 1 MDRSENIRLTMRSRLVSL.....RASFGVNPATPTADTYLQX 957

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3373863 seqs, 2124095041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame_plus_p2n.model -DRV=xlh
-Q=/cgn2_1/USPTO.spool/US09917376/runat_14052004_085256_4609/app_query.fasta_1.1095
-DB=N Geneseq 250an04 -QFMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=biosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09917376.@CGN_1_1_581.@runat_14052004_085256_4609 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq 29Jan04:.*
1: Geneseqn1980s:.*
2: Geneseqn1990s:.*
3: Geneseqn2000s:.*
4: Geneseqn2001as:.*
5: Geneseqn2001bs:.*
6: Geneseqn2002s:.*
7: Geneseqn2003as:.*
8: Geneseqn2003bs:.*
9: Geneseqn2003cs:.*
10: Geneseqn2004s:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5134	100.0	2869	7 ABZ77632	Nucleotid
2	5134	100.0	2869	ADD22922	Acidother
3	2174	42.3	2823	7 ABV76941	Nucleotid
4	1558.5	30.4	2849	9 ADD42060	Trichoder
5	1167.5	22.7	3668	9 ADD24893	DNA encod
6	940	18.3	5698	9 ADD42054	Trichoder
7	823	16.0	3687	7 ABZ77634	Nucleotid
8	702	13.7	2289	7 ABZ77633	Nucleotid

9	429	8.4	3365	7 ABZ776162	Abz776162 A. cellul
10	409	8.0	11706	2 AAX55661	Aax55661 DNA seque
11	409	8.0	11707	6 AAD26525	Aad26525 Active ce
12	392.5	7.6	1103	3 AAF14988	Aaf14988 Trichoder
13	365.5	7.1	651	5 AAS84956	Aas84956 DNA encod
14	345	6.7	18596	4 AAF31109	Aaf31109 Thymidyla
15	345	6.7	18596	5 AAC91215	Aac91215 Human thy
16	345	6.7	18596	6 ABL67927	AbL67927 Ovary can
17	345	6.7	18596	6 ABL62854	AbL62854 Breast ca
18	345	6.7	18596	6 ABL63078	AbL63078 Breast ca
19	345	6.7	18596	6 ABK43334	Abk43334 Human Thy
20	345	6.7	18596	6 ABN95092	Abn95092 Gene #159
21	334.5	6.5	6415	2 AAX55662	Aax55662 DNA seque
22	334	6.5	6416	6 AAD26526	Aad26526 Active ce
23	330.5	6.4	7407	7 ACA42281	Aca42281 Prokaryot
24	330.5	6.4	7407	8 ACC59398	Acc59398 Microbial
25	323	6.3	2600	2 AAQ15178	Aaq15178 Portion o
26	318	6.2	1080	6 ABI99537	Abi99537 Mouse isc
27	302	5.9	1050	4 AAD21685	Aad21685 Mutationa
28	302	5.9	2849	4 AAD21684	Aad21684 Human ret
29	296.5	5.8	4767	3 AAA14667	Aaa14667 Nucleotid
30	293	5.7	4818	3 AAA14669	Aaa14669 Nucleotid
31	290	5.6	1232	4 AAD11124	Aad11124 Human sma
32	289	5.6	324	7 ACA25426	Aca25426 Prokaryot
33	288.5	5.6	4571	3 AAA14662	Aaa14662 Nucleotid
34	288	5.6	288	6 ABL51804	AbL51804 Hydroxypr
35	285.5	5.6	5184	8 ACC80175	Acc80175 Human MC2
36	282.5	5.5	9903	7 ACA40315	Aca40315 Prokaryot
37	282.5	5.5	110000	4 AA199682_04	Continuation (5 of
38	281.5	5.5	432	4 AA156352	AA156352 Probe #12
39	281.5	5.5	518	4 AA143371	AA143371 Probe #12
40	281.5	5.5	110000	6 AA199683_37	Continuation (38 o
41	281.5	5.5	110000	6 ABS55320_0	ABS55320 Human tra
42	281.5	5.5	341511	6 ABS55200	ABS55200 Genomic D
43	280.5	5.5	3008	7 AAD55730	Aad55730 Argiopoic
44	280	5.5	49999	2 AAZ23895	Aaz23895 Murine LO
45	279	5.4	4547	3 AAA14664	Aaa14664 Nucleotid

ALIGNMENTS

RESULT 1
ABZ77632
ID ABZ77632 standard; DNA; 2869 BP.
XX ABZ77632;
AC ABZ77632;
XX
DT 03-JUN-2003 (first entry)
XX
DE Nucleotide sequence of the avicelase AvIII.
XX
KW Avicelase; AvIII; glycoside hydrolase; enzyme; cellulase; biofuel;
KW detergent; pulp processing; paper processing; feed processing; textile;
KW cellulose; gene; ss.
XX
OS Acidothermus cellulolyticus.
XX
FH Key Location/Qualifiers
CDS 1..2869
FT /*tag= a
FT /partial
FT /product= "AvIII"
FT /transl_except= (pos:2869,aa:Xaa)
FT /note= "Xaa is an unspecified residue"

PA (WIDE) MIDWEST RES INST.
 XX Ding S, Adney WS, Vinzant TB, Himmel MB;
 XX WPI; 2003-248177/24.
 DR P-PSDB; ABP73015.
 XX New thermostable Aviii peptide from *Acidothermus cellulolyticus*, useful
 PT for degradation of cellulose or in generating anti-Aviii antibodies for
 PT purifying recombinant Aviii polypeptides from genetically engineered
 PT host cells.
 XX
 XX Claim 3; Page 24; 44pp; English.
 XX
 XX The present sequence encodes a thermostable avicelase polypeptide,
 CC designated Aviii. Aviii is a member of the glycoside hydrolase family
 CC of enzymes, and is a cellulase. Aviii is useful in the conversion of
 CC biomass to biofuels and biofuel additives. It may be useful in the
 CC production of detergents, pulp and paper processing, food and feed
 CC processing and in textile processes. The thermostable Aviii peptide is
 CC useful in the degradation of cellulose, and in generating specific anti-
 CC Aviii antibodies that are useful in purifying recombinant Aviii
 CC polypeptides from genetically engineered host cells, in detecting Aviii
 CC polypeptide expression, as well as a reagent tool for characterizing the
 CC molecular actions of the polypeptide. The Aviii polynucleotide is useful
 CC as a source of probes or primers in various diagnostic assays
 XX
 XX Sequence 2869 BP; 546 A; 857 C; 900 G; 565 T; 0 U; 1 Other;

Alignment Scores:

Pred. No.: 1,12e-172 Length: 2869
 Score: 5134.00 Matches: 956
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 99.98% Indels: 0
 DB: 7 Gaps: 0

US-09-917-376-1 (1-957) x ABZ77632 (1-2869)

QY 1 MetAspArgSerGluAsnIleArgLeuThrMetArgSerArgLeuValSerLeuLeu 20
 DB 1 ATGGATCGTTGGAGAACATCCGCTGACTGATGAGATCAGACGATGGTATCACTGCTC 60
 QY 21 AlaAlaThrAlaSerPheAlaValAlaAlaAlaLeuGlyValLeuProfileAlaIleThr 40
 DB 61 GCCGCCACTCGCTTCGCGGTGGCGCGCTCTGGAGTCTGCCCATCGGATACG 120
 QY 41 AlaSerProAlaHisAlaAlaThrThrGlnProTyrThrTrpSerAsnValAlaIleGly 60
 DB 121 GCTTCTCTCGGCACGCGCGAGACTCAGCCGTACACCTGGAGCAACGTTGGCGATCGGG 180
 QY 61 GlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrVal 80
 DB 181 GCGCGCGCGCTTTGTGACGGGATCGTCTTCAATGAAGGTGACCGGAATCTGTACGTG 240
 QY 81 ArgThrAspIleGlyMetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeu 100
 DB 241 CGACGGACATCGGGGGATGATCGATGGATGCCCAACGGCGGTGGATCCCTCTT 300
 QY 101 LeuAspTrpValGlyTrpAsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAsp 120
 DB 301 CTGGATTTGGTGGATGGAACAAATGGGGGTCAACCGCGCTCGTCAGCATTTGGCGAGAC 360
 QY 121 ProIleAsnThrAsnIleValTrpAlaValAlaGlyMetTyrThrAsnSerTrpAspPro 140
 DB 361 CCAGTCAATCACTAACAGGTATGGCGCGCTCGGAATGTACACCAACAGCTGGGACCCA 420
 QY 141 AsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeu 160
 DB 421 AACACGGACGATTTCTCCGCTCTGTCTGATCAGGGCGCAACGTGGCAATAACGCCCTG 480
 QY 161 ProPheIleLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAsp 180

DB 481 CCGTTCAAGCTTTGGCGCAACATCCCGGGCGTGGATGGAGCGGCTTGGGTGGAT 540
 QY 181 ProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSer 200
 DB 541 CCAAAACAATGACAACATTTCTGTAATTTGGCGCCCCGAGCGGCAAGGGCTCTGGAGAAGC 600
 QY 201 ThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyrIle 220
 DB 601 ACAGATTCCCGCGGACCTGTGTCAGATGACGAACATTTCCGGACGTAGGACGTACATT 660
 QY 221 AlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAla 240
 DB 561 GCAAAATCCCACTGACACGCGGCTATCAGACGATATTCAGGGCGTCTGTGGGTGGCT 720
 QY 241 PheAspIleSerSerSerSerLeuGlyGlnIleAsnIleThrIlePheValGlyValAla 260
 DB 721 TTCGACAAAGTCTTCGTATCTCGTCCGCAAGCGAGTAGAAGACCATTTTGTGGCGTGGCG 780
 QY 261 AspProAsnAsnProValPheTrpSerArgAspGlyGlyAlaThrTrpGlnIleValPro 280
 DB 781 GATCCCAATAATCCGGTCTTCTGGAGCAGACGCGCGCGACGTGGCGAGCGGTGGCG 840
 QY 281 GlyAlaProThrGlyPheIleProHisIleValPheAspProValAsnHisValLeu 300
 DB 841 GTGCGCGCACCGGCTTCATCCGCAAGGGCGTCTTTGACCCGGTCAACACGTCCTC 900
 QY 301 TyrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLys 320
 DB 901 TATATTGCCACCAGCAATACGGGTGGTCCGTATGACGGGAGCTCCGGCGACGTCTGAAA 960
 QY 321 PheSerValThrSerGlyThrTrpThrArgIleSerProValProSerThrAspThrAla 340
 DB 961 TTTCTCGGTGACCTCCGGGACATGACGCGGAATCAGCCCGGTACCTTCGACGACACGCGC 1020
 QY 341 AsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMet 360
 DB 1021 AACGACTACTTTGGTTACAGCGGCTCTACTATCGACCCCGACACCCGCAACAGATAATG 1080
 QY 361 ValAlaThrGlnIleSerTrpProAspThrIleIlePheArgSerThrAspGlyGly 380
 DB 1081 GTGGCAACCCAGATATCGTGGTGGCGGACACCAATAATCTTTCGGAGCACCGCGCGGT 1140
 QY 381 AlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrVal 400
 DB 1141 CGCAGCTGGACGCGATCTGGGATTCGAGATTATCCCAATCGAAGCTTGGATATGTG 1200
 QY 401 LeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnProProValPro 420
 DB 1201 CTTGACATTTCCGGGAGGCTTGGCTGACCTTCGGCGTACAGCCGAATCTCCCGTACCC 1260
 QY 421 SerProIleLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArg 440
 DB 1261 AGTCGAGCTCGGCTGGATGGATGAAGGATGGCAATCGATCCGTCAACTCTGATCGG 1320
 QY 441 MetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrIleTrpAspSer 460
 DB 1321 ATGCTCTACGGAACAGGCGCGACGTTGTACGCAACAATGATCTCAGAAATGGGACTCC 1380
 QY 461 GlyGlyGlnIleHisIleAlaProMetValIleGlyLeuGluGluThrAlaValAsnAsp 480
 DB 1381 GCGCGCCAGATTCATATCGCGCGATGGTCAAAAGATTGGAGGACGCGCGTAAACGAT 1440
 QY 481 LeuIleSerProSerGlyValAlaProLeuIleSerAlaLeuGlyAspLeuGlyPhe 500
 DB 1441 CTCATCAGCCCGCTCTGGCGCCCCGCTCATCAGCGCTCTCGAGACCTCGCGGCTTC 1500
 QY 501 ThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProValPheThrThr 520
 DB 1501 ACCCAGCGCGAGTTACTTCGCGTGCATCGACGATCTTCACGTCACCGGTGTTACGACC 1560
 QY 521 GlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPhe 540
 DB 1561 GGCACACGCGTCCGATATCGCGAATTGAATCCGTCGATCATCGTTCCGCGCTGGAAGTTTC 1620

QY 541 AspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsn 560
DB 1621 GATCCATCGAGCGCAACCGCAACGACGCGTCCGCTTCTCGACGACGCGCGCAAGAAC 1680
QY 561 TrpPheGlnGlySerGluProGlyGlyValThrThrGlyGlyThrValAlaAlaSerAla 580
DB 1681 TGGTTCCAAAGGACGCAACCTGGCGGGGTGACACGCGGGCGCAACCTCGCGCATCGGCC 1740
QY 581 AspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValValAlaVal 600
DB 1741 GACGGCTCTCGTTTCGTCTGGGCTCCCGCGCATCCCGGTGAGCTGTGTGTACGAGTC 1800
QY 601 GlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSer 620
DB 1801 GGATTGGCAACTCTCTGGGCTGCTTCGCAAGGTGTTCCGCGCAATCCCGCAGATCCGCTCA 1860
QY 621 AspArgValAsnProIysThrPheThrAlaLeuSerAsnGlyThrPheThrArgSerThr 640
DB 1861 GACCGGGTGAATCCAAAGACTTTCTATGCCCTATCCAAATGGAACTTCTATCGAAGCAG 1920
QY 641 AspGlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGly 660
DB 1921 GACCGCGCGTGACATTCACACGCTCGCGCGCGTCTTCGAGACGCGGTCCGCTCGGT 1980
QY 661 ValMetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGly 680
DB 1981 GTCATGTTCACGCGGTGCTCGAANAAGAGCGATCTGTGGCTCGCTCATCGAGCGGG 2040
QY 681 LeuThrHisSerThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAla 700
DB 2041 CTTTACCACTCAACCAATGGCGGACGAGTGTGCTGCAATCACCGCGGTATCTCCGCG 2100
QY 701 ValAsnValGlyPheGlyLysSerAlaProGlySerSerTrpProAlaValPheValVal 720
DB 2101 GTGAACGTGGGATTTGGTAAGTCTCGCGCCGGGTGCTATACCCAGCCGTCTTTGTGCTC 2160
QY 721 GlyThrIleGlyValThrGlyAlaTyrArgSerAspAspCysGlyThrTrpVal 740
DB 2161 GGCACGATCGAGGGCTTACGGGGGTACCGCTCCGACGACTGTGGGACGACCTGGGTA 2220
QY 741 LeuIleAsnAspAspGlnHisGlnTrpGlyAsnTrpGlyAlaIleThrGlyAspHis 760
DB 2221 CTGATCAATGATGACGACCAATACGGAATTTGGGCAAGCAATCACCGGTGACCAC 2280
QY 761 AlaAsnLeuArgArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIle 780
DB 2281 GCGAATTTACGCGGGTGTACATAGCACGAAACGCGCGTGGAAATGTATACGGGGACATT 2340
QY 781 GlyGlyAlaProSerGlySerProSerProSerValSerProSerAlaSerProSerLeu 800
DB 2341 GGTGTGCGCCGCTCGGATCGCGCTCCGTCGTGGTGAGTCCGTGGCTTTGGCGAGCCCTG 2400
QY 801 SerProSerProSerProSerProSerProSerProSerProSerProSerProSerPro 820
DB 2401 AGCCCGAGCCCGAGCGAGCTCGCCATCGCGTCCGCTCGCGGAGCTCGAGTCCA 2460
QY 821 SerSerSerProSerProSerProSerProSerProSerProSerProSerProSerPro 840
DB 2461 TCTCTGTCGCGCTCTCCGTCGCGCTCACCATCGCCGAGTCCGTCGCTCGCTCCGTCACCA 2520
QY 841 SerAlaSerProSerProSerProSerProSerProSerProSerProSerProSerPro 860
DB 2521 TCGGCTGCGCCGAGCCGCTCTCGTCACCGAGCCGCTCTCTGTCACCGCTCTTCGTCGCG 2580
QY 861 SerProThrProSerSerProValSerGlyValValValValValValValValValVal 880
DB 2581 AGCCCAACCGCTCGCTCGCGGGTGTGGGTGGGGTGAAGGTGAGTATAGATATAT 2640
QY 881 AspSerAlaProGlyAspAsnGlnIleLysProGlyLeuGlnValValAsnThrGlySer 900
DB 2641 GATTTCGCGCGCGGTGATTAATCAGATCAAGCGCGGTTTTCAGGTGTGTGAATACCGGGT 2700

QY 901 SerSerValAspLeuSerThrValThrValArgTyrTrpPheThrArgAspGlyGlySer 920
DB 2701 TCCTCGGTGATTTGTTCGACGGTGACGGTGGGTACTGGTTTACCCGGGATGGTGGCTCG 2760
QY 921 SerThrLeuValTyrAsnCysAspTrpAlaAlaIleGlyCysGlyAsnIleArgAlaSer 940
DB 2761 TCACACTGGTGTACAACTGTGACTGGCGGCGATCGGGTGTGGGAATATCCGCGCTCG 2820
QY 941 PheGlySerValAsnProAlaThrProThrAlaAlaAspThrTyrLeuGln 956
DB 2821 TTCGGCTCGTGAACCGCGGACGCCGACGCGGACACCTACTACCTGCAG 2868
RESULT 2
ADD22922
ID ADD22922 standard; DNA; 2869 BP.
XX
AC ADD22922;
XX
DT 15-JAN-2004 (first entry)
XX
DE Acidothermus cellulolyticus avicelase Aviii DNA.
XX
KW Aviii; cellulose reduction; agricultural biomass; municipal solid waste;
KW glycoside hydrolase; avicelase; ds; gene.
XX
OS Acidothermus cellulolyticus.
XX
FH Key Location/Qualifiers
FT CDS 1..2868
FT /*tag= a
FT /product= "Aviii"
XX
PN US2003108988-A1.
XX
PD 12-JUN-2003.
XX
PF 18-OCT-2002; 2002US-00155400.
XX
PR 28-JUL-2001; 2001US-00917376.
XX
PA (DING/) DING S.
PA (ADNE/) ADNEY W S.
PA (VINZ/) VINZANT T B.
PA (HIMM/) HIMMEL M E.
XX
PI Ding S, Adney WS, Vinzant TB, Himmel ME;
XX
DR WPI; 2003-810853/76.
DR P-PSDB; ADD22921.
XX
PT New isolated thermal tolerant avicelase polynucleotide useful for
PT detection of a polynucleotide encoding Aviii and for reducing cellulose
PT in a starting material, e.g. municipal solid waste.
XX
XX Claim 17; SEQ ID NO 2; 29pp; English.
XX
CC The invention relates to an isolated polynucleotide molecule encoding a
CC thermostable Aviii polypeptide. The polynucleotide is useful for
CC detection of a polynucleotide encoding Aviii. The polynucleotide is
CC useful for reducing cellulose in a starting material which involves
CC administering to the starting material, e.g. agricultural biomass or
CC municipal solid waste, a polypeptide molecule of the polynucleotide. The
CC method further comprises administering a second polypeptide molecule
CC chosen from the glycoside hydrolase family of proteins. The present
CC sequence represents DNA encoding the Acidothermus cellulolyticus
CC avicelase Aviii.
XX
SQ Sequence 2869 BP; 546 A; 857 C; 900 G; 565 T; 0 U; 1 Other;

Alignment Scores:

Pred. No.: 1.12e-172 Length: 2869
Score: 5134.00 Matches: 956
Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0			
Query Match: 99.98% Indels: 0			
DB: 5 Gaps: 0			
US-09-917-376-1 (1-957) x ADD22922 (1-2869)			
Qy	1	MetAspArgSerGluAsnIleArgLeuThrMetArgSerArgArgLeuValSerLeuLeu	20
Db	1	ATGGATCGTTCGGAGAACATCCGCTCTCACTATGAGATCACGACGATTGGTATCACTGCTC	60
Qy	21	AlaAlaThrAlaSerPheAlaValAlaAlaLeuGlyValLeuProIleAlaIleThr	40
Db	61	CGCCCACTCGCTGCTTCGGCGTGGCGCGCTCTGGAGTTCTGCCCATCGCGATACG	120
Qy	41	AlaSerProAlaHisAlaAlaThrThrGlnProTyrThrTyrSerAsnValAlaIleGly	60
Db	121	GCCTTCTCTGGCAGCGCGGACGACTCAGCGGTACACCTGGAGCAACGTTGGGATCGGG	180
Qy	61	GlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrVal	80
Db	181	GGCGGGCGGCTTGTTCGACGGGATCGTCTTCAATGAAGGTGCACCGGGAATTTGTACGTG	240
Qy	81	ArgThrAspIleGlyMetTyrArgTyrAspAlaAlaAsnGlyArgTyrIleProLeu	100
Db	241	CGGACGACATCGGGGGGATGTATCGATGGATGCCGCCAACGGCGGTGGATCCCTCTT	300
Qy	101	LeuAspTyrValGlyTyrAsnAsnTyrGlyTyrAsnGlyValValSerIleAlaAlaAsp	120
Db	301	CTGGATGGGTGGATGGACAAATTTGGGGTACAAACGGCGTCTGACGATTCGGGACAG	360
Qy	121	ProIleAsnThrAsnLysValTyrAlaAlaValGlyMetTyrThrAsnSerTyrAspPro	140
Db	361	CCGATCAATTAACAAGGTATGGCGCGCTCGGAATGTACCAACAGCTGGGACCCA	420
Qy	141	AsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTyrGlnIleThrProLeu	160
Db	421	AACGACGGAGCGATTCCTCGCTCGTCTGATCAGGGCGCAACGTTGGCAATAAAGCCCTG	480
Qy	161	ProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAsp	180
Db	481	CCGTTCAAGCTTGGCGGCAACATGCCCGGGGTGGAAATGGCGAGCGGCTTGGGTGGAT	540
Qy	181	ProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTyrArgSer	200
Db	541	CCAAACAATGACAAATTCCTGATTTTCGGCGCCCGAGCGCAAAAGGCTCTGGAGAAGC	600
Qy	201	ThrAspSerGlyAlaThrTyrSerGlnMetThrAsnPheProAspValGlyThrTyrIle	220
Db	601	ACAGATTCCGGCGGACCTGCTCCAGATGACCAACTTTCCGGACGTAGGACGATACATT	660
Qy	221	AlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTyrValAla	240
Db	661	GCAATCCCACTGACACGACGGCTATCAGACGATATTCAAGGCGTCTGCTGGGTGCTT	720
Qy	241	PheAspLysSerSerSerLeuGlyGlnAlaSerLysThrIlePheValGlyValAla	260
Db	721	TTTCGACAAGTCTCTCATCGCTCGGCAAGCGAGTAAGACCAATTTTGTGGGCGTGGCG	780
Qy	261	AspProAsnAsnProValPheThrTyrSerArgAspGlyGlyAlaThrTyrGlnAlaValPro	280
Db	781	GATCCCAATAATCCGGTCTTCTGGAGCAGAGACGGCGCGCGACGTCGGCAGGCGGTCCG	840
Qy	281	GlyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsnHisValLeu	300
Db	841	GGTGGCGGACCGGCTTTCATCCCGCAACAGGGCGTCTTTACCGCGGTCAACACGCTGCTC	900
Qy	301	TyrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspValTyrLys	320
Db	901	TATATTGCCACCAACATACGGGTGTCCTGTATGACGGGAGTCCGCGGACGCTCGAAA	960
Qy	321	PheSerValThrSerGlyThrTyrThrArgIleSerProValProSerThrAspThrAla	340
Db	961	TTCTCGGTGACCTCCGGGACATGACGCGAATCAGCCCGGTACCTTCGACGGACACGGCC	1020
Qy	341	AsnAspTyrPheGlyTyrSerGlyLeuThrThrIleAspArgGlnHisProAsnThrIleMet	360
Db	1021	AACGACTACTTTGGTTACAGCGGCTCACTATCGACCGCCAGCACCCGAAACGATAATG	1080
Qy	361	ValAlaThrGlnIleSerTyrTyrProAspThrIleIlePheArgSerThrAspGlyGly	380
Db	1081	GTGGCAACCCAGATATCGTGTGGCGGACACCAATAATCTTTTCGGAGCACCCGACGGGT	1140
Qy	381	AlaThrTyrThrArgIleTyrAspTyrThrSerTyrProAsnArgSerLeuArgTyrVal	400
Db	1141	CGGACGTGGACCGGATCTGGGATTCGACGAGTTATCCCAATCGAAGCTTCGATATGTG	1200
Qy	401	LeuAspIleSerAlaGluProTyrPheGlyValGlnProAsnProProValPro	420
Db	1201	CTTGACATTTCCGGGAGCCTTGGCTGACCTTCGGCGTACAGCCGAATCTCTCCGTAACC	1260
Qy	421	SerProLysLeuGlyTyrMetAspGluAlaMetAlaIleAspProPheAsnSerAspArg	440
Db	1261	AGTCCGAAGCTCGGCTGGATGATGAAGCATGGCAATCGATCCGTTCAACTCTCATCGG	1320
Qy	441	MetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTyrAspSer	460
Db	1321	ATGCTCTACGGAACAGGCGCGCGTGTTCACCAACAATGATCTCAGAAAGTGGGACTCC	1380
Qy	461	GlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluGluThrAlaValAsnAsp	480
Db	1381	GGCGGCGAGATTATATCGCGCGATGTTCAAGATTGGAGAGACGCGCGTAAACGAT	1440
Qy	481	LeuIleSerProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyPhe	500
Db	1441	CTCATCAGCCCGCGCTCTGGCGCCCGCTCATCAGCGCTCTCGGAGACCTCGCGCGCTTC	1500
Qy	501	ThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProValPheThrThr	520
Db	1501	ACCACGCGGAGCTTACTGCGGCTGCGCATCGAGATCTTCACGTACCGGTGTTACGACC	1560
Qy	521	GlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPhe	540
Db	1561	GGCACGAGCTCGACTATCGGAATTGAATCCGTCGATCATCGTTCGCGTGGAAAGTTTC	1620
Qy	541	AspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyLysAsn	560
Db	1621	GATCCATCGAGCAACCCGACAGCAGAGCGCGCTCGCGTCTTCGACACAGCGCGCGCAAGAC	1680
Qy	561	TrpPheGlnGlySerGluProGlyValThrThrGlyGlyThrValAlaAlaSerAla	580
Db	1681	TGGTTCCAAAGCAGCAACCTGGCGGGGTGACGACGGGCGGCACCGTCCGCCATCGGCC	1740
Qy	581	AspGlySerArgPheValTyrAlaProGlyAspProGlyGlnProValValTyrAlaVal	600
Db	1741	GACGCTCTCGTTTCGTCCTGGGCTCCCGCGCATCCCGGTACGCTGCTGGTGTACGACGTC	1800
Qy	601	GlyPheGlyAsnSerTyrAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSer	620
Db	1801	GGATTTGGCAACTCTCGGCTGCTTCGCAAGGTTCGCCGCAATGCCAGATCCGCTCA	1860
Qy	621	AspArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThr	640
Db	1861	GACCGGTTGAATCCAAAGACTTTCTATGCCCTATCCCAATGGAAACCTTCTATCGAAGCAG	1920
Qy	641	AspGlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGly	660
Db	1921	GACGGCGGCTGACATTCCAACCGGTCGCGCGGCTCTTCGAGGAGCGGTCGCTCGGT	1980
Qy	661	ValMetPheHisAlaValProGlyLysGlyAspLeuTyrLeuAlaAlaSerSerGly	680
Db	1981	GTGATTTCCACGCGGTCCTGGAAAGAGCGATCTGTGGCTCGTGCATCGACGCGG	2040
Qy	681	LeuTyrHisSerThrAsnGlyGlySerSerTyrSerAlaIleThrGlyValSerSerAla	700
Db	2041	CTTTACCACCTAACCAATGGCGGACGAGCTTGGTCTGCAATCACCGGGGTATCTCTCCGG	2100

```

QY 701 ValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheValVal 720
Db 2101 GTGAACGTGGGATTTGGTAAGTCTGCGCCCGGGTCTGCATACCCAGCGCTTTGTGCTC 2160
QY 721 GlyThrIleGlyGlyValThrGlyValTyrArgSerAspAspCysGlyThrThrVal 740
Db 2161 GGCACGATCGAGCGGTTCAGGGGCGGTACCGCTCCGACGACTGTGGACGACCTGGGTA 2220
QY 741 LeuIleAsnAspAspGlnHisGlnTyrGlyAsnTyrGlyGlnAlaIleThrGlyAspHis 760
Db 2221 CTGATCAATGATGACACGACCAATACGAAATTTGGGACAAATCACCAGGTGACCA 2280
QY 761 AlaAsnLeuArgArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIle 780
Db 2281 GCGAATTTACGGCGGTGTACATAGGCACCAACGCGCGTGGAAATGTATACCGGGACATT 2340
QY 781 GlyGlyAlaProSerGlySerProSerProSerProSerProSerProSerProSerLeu 800
Db 2341 GGTGGTGGCGCGTCCGATCGCGTCTCCGTCGGTGAAGTCGCTCGGCTTCGCGAGCCTG 2400
QY 801 SerProSerProSerProSerProSerProSerProSerProSerProSerProSerPro 820
Db 2401 AGCCGAGCGCGAGCCGACGAGCTCCCATCGCCGTCGCGCTCGCGAGCTCGAGTCCA 2460
QY 821 SerSerSerProSerProSerProSerProSerProSerProSerProSerProSerPro 840
Db 2461 TCCGTCGCGCGTCTCGCTCGCGTCAACATCGCGAGTCCGTCGCTCGCTCGCTCA 2520
QY 841 SerAlaSerProSerProSerProSerProSerProSerProSerProSerProSerPro 860
Db 2521 TCGCGCTCGCGAGCGGTCTTCGTCAACGAGCCGCTTCGTCACCGCTCTTCGTCGCG 2580
QY 861 SerProThrProSerProSerProSerProValSerGlyGlyValValGlnTyrLysAsn 880
Db 2581 AGCCCAACGCGCTCGCTCGCGCGTCTCGGCTGGGTGAAGTGAAGTGAAGTGAAGT 2640
QY 881 AspSerAlaProGlyAspAsnGlnIleLysProGlyLeuGlnValValAsnThrGlySer 900
Db 2641 GATTTCGGCGCGGTGATATATCATGATCATGACCGCGGTTTGCAGTGGTGAATACCG 2700
QY 901 SerSerValAspLeuSerThrValThrValArgTyrTrpPheThrArgAspGlyGlySer 920
Db 2701 TCGTCGGTGGATTTGTTCGAGCGTTCGAGCGTTCGCTGCTGCTGCTGCTGCTGCTG 2760
QY 921 SerThrLeuValTyrAsnCysAspTrpAlaIleGlyCysGlyAsnIleArgAlaSer 940
Db 2761 TCGACATGTTGTATCAACTGTGACTCGCGCGCGATCGCGTGGGAATATCCGCGCTCG 2820
QY 941 PheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 956
Db 2821 TTCGGCTCGGTGAACCGCGGACGCCGACCGCGGACACCTACCTGCTGAG 2868

RESULT 3
ABV76941
ID ABV76941 standard; DNA; 2823 BP.
XX AC ABV76941;
XX DT 03-MAR-2003 (first entry)
XX DE Nucleotide sequence of a xyloglucanase enzyme.
XX KW Xyloglucanase; enzyme; family 74; glycosyl hydrolase; cellulosic fiber;
XX OS Jonesia sp.
XX PH Key Location/Qualifiers
FT CDS 1..2823
FT /tag= a
FT /product= "xyloglucanase enzyme"
FT sig_peptide 1..69
FT /tag= b

```

```

mat_peptide 70..2820
/*tag= c
WO200277242-A2.
03-OCT-2002.
27-MAR-2002; 2002MO-DK000210.
27-MAR-2001; 2001DK-00000504.
(NOVO ) NOVOZYMES AS.
Duffner F, Sjolholm C;
P-PSDB; ABB99489.
WPI; 2003-092855/08.
New xyloglucanase enzyme belonging to family 74 of glycosyl hydrolases,
and endogenous to a bacterium, useful in the textile industry for
improving properties of cellulosic fibers, yarn or fabric.
Claim 11; Page 68-72; 76pp; English.
The present sequence encodes a xyloglucanase enzyme, belonging to family
74 of glycosyl hydrolases. The enzyme is isolated from Jonesia sp.
DSM14140. The enzyme is useful in processes for machine treatment of
fabrics. It is also useful in the textile industry for improving the
properties of cellulosic fibers, yarn, woven or non-woven fabric, and in
a textile scouring process step. The xyloglucanase enzyme is also useful
in the cellulose fiber processing industry for rating of fibers such as
hemp, jute, flax and linen. It is useful for preventing binding of
certain soils to the xyloglucan left on the cellulosic material
XX SQ Sequence 2823 BP; 665 A; 803 C; 764 G; 591 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1,99e-68 Length: 2823
Score: 2174.00 Matches: 430
Percent Similarity: 59.48% Conservative: 141
Best Local Similarity: 44.79% Mismatches: 311
Query Match: 42.34% Indels: 78
DB: 7 Gaps: 16

US-09-917-376-1 (1-957) x ABV76941 (1-2823)
QY 10 ThrMetArgSerArgArgLeuValSerLeuLeuAlaAlaThrAlaSerPheAlaValAla 29
Db 13 ACCGTGGTGGCCACAGCAGCTAACTGCGTCTCGCCACGACTCGCGTGGCATCGCGCC 72
QY 30 AlaAlaLeuGlyValLeuProIleAlaIleThrAlaSerProAlaHiAlaAlaThrThr 49
Db 73 ACATCA-----ACCCCATCGCTAGTAATCGACCGCCCTCTTTCATCATCACT 123
QY 50 GlnProTyrThrTrpSerAsnValAlaIleGlyGlyGlyGlyPheValAspGlyIleVal 69
Db 124 TCCGGCTATAGCTGTCGACGTTGAGATTGTGGCGCGCGGTTTTGTCCCTGGAATTGTC 183
QY 70 PheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGlyMetTyrArg 89
Db 184 TTTAACCCAGAAAGACCCAGGCTTGGTGTATGCCGACAGATATTTGGTGGACGCTACCGG 243
QY 90 TrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrp 109
Db 244 CTCAATGACTCCACCGCGGTGGATCCCTCTTACGGACCATATTTGGTTGGGATGACTGG 303
QY 110 GlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnIleValTrpAla 129
Db 304 AGCCACTCCGGAATCTGTCTCTGCGCTGACCTGACCTGTGTACACCAACCGCGTGACCTT 363
QY 130 AlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSer 149
Db 364 GCAGCAGCAGTACTCCATGGACTGGATCCAAACGAGGTGCTATTATTGGTTCCGCG 423

```



```

Db      2574  GTCCGG 2580
RESULT 5
ADD24893
ID      ADD24893 standard; DNA; 3668 BP.
XX
AC      ADD24893;
XX
DT      15-JAN-2004 (first entry)
XX
DE      DNA encoding Xanthomonas campestris cellulase #1.
XX
KW      Directed genetic engineering; galactomannanase; reduced activity;
KW      enhanced activity; xanthan gum production; suspension stability;
KW      emulsion stability; temperature resistance; pseudoplasticity; amylase;
KW      cellulase; extracellular protease; intracellular protease;
KW      glucose dehydrogenase; gene; ds.
XX
OS      Xanthomonas campestris.
XX
PN      US2003036176-A1.
XX
PD      20-FEB-2003.
XX
PF      10-AUG-2001; 2001US-00927827.
XX
PR      28-MAR-2001; 2001US-0279493P.
XX
PA      (BOWE/) BOWER S G.
PA      (RAMS/) RAMSEIER T M.
XX
PI      Bower SG, Ramseier TW;
XX
DR      WPI; 2003-625389/59.
XX
P-PSDB; ADD24919.
XX
New transformed cell or organism having reduced or enhanced activity of
PT      at least one protein, useful for producing xanthan gum, which are useful
PT      for providing formulations and properties.
XX
Claim 15; SEQ ID NO 21; 135pp; English.
XX
The present invention relates to polypeptide and polynucleotide sequences
CC      from Xanthomonas campestris which may be used for activity reduction or
CC      enhancement using directed genetic engineering. A transformed cell or
CC      organism having reduced or enhanced activity of at least one such protein
CC      e.g. galactomannanase can be generated by disrupting the gene encoding
CC      the protein. The activity of the protein is reduced by the presence of an
CC      antisense nucleic acid sequence. The nucleic acid sequence of the gene
CC      encoding the protein is a recombinant sequence having at least one
CC      mutation as compared to the wild-type gene encoding the protein. The
CC      transgenic cell or microorganism are useful for producing xanthan gum,
CC      which are useful for providing formulations and properties, such as long-
CC      term suspension and emulsion stability in alkaline, acid, and salt
CC      solutions, temperature resistance, and pseudoplasticity. The present
CC      sequence encodes an enzyme relating to the present invention.
XX
SQ      Sequence 3668 BP; 570 A; 1181 C; 1292 G; 625 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:      7e-33      Length:      3668
Score:          1167.50
Percent Similarity: 48.55%
Best Local Similarity: 33.30%
Query Match:    22.74%
DB:            9
Gaps:          28

US-09-917-376-1 (1-957) x ADD24893 (1-3668)

Qy      23 ThrAlaSerPheAlaValAlaAlaLeuGlyValLeuProfilAlaThrAlaSer 42
Db      1136 ACCAACAGGAGCATGGCTGGCTGCTTGGAGCCTGCTG---CTGTTGCTCTTTTCGACG 1192

```


399 TyrValLeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAnProPro 418
2231 GCGGTGTTGATCACAGCGCGCGCGTGG-ACCGCACATGCCACGCGCGCAC- 2281
419 ValProSerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAnSer 438
2282 -TGATGGGG- -GCGCTGGCGATCGATCCGTTTCGACGGC 2317
439 AspArgMetLeuTyrGlyThrGlyAlaThrLeuThrAlaThrAsnAspLeuThrLysTrp 458
2318 AACCATGCGCTGTTCGACCGCTACCGCATCTGGCCCTCGGCCATCTG- 2368
459 AspSerGlyGlyGlnIleHisIleAlaPro- -MetVal 470
2369 -CAGGATTTGCGCGCACCGCGCGCGCTGTCAGTGGTGTTCAGGAC 2416
471 LysGlyLeuGluGluThrAlaValAsnAspLeuIleSerProSerGlyAlaProLeu 490
2417 CGTGGGCTGGAAACCGCGTGGCGCTGGACCTGCTCAGCCCGATGCTGGCGCGCATCTG 2476
491 IleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAlaValProSer 510
2477 CTCAGCGCGCTCGCGCATATCGACGGTTCGCGCATGACGACCTGGACCGCTG- 2533
511 ThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyrAlaGluLeuAsn 530
2534 TTGCACTAGCGCGGCCACCGCTGACCAATGGCGAAAGCATGATGCGCGCGCCAGCGC 2593
531 ProSerIleValAlaGlySerPhe-AppProSerSerGlnProAnAspArgHi 550
2594 CCGCAGTGGGTGGTGGCGACGCTACCGTGGCGCGCGCGCGCGCAACGAAATCCGCGCG 2653
550 sValAlaPheSerThrAspGlyGlyLysAsnTrpPheGln- -GlySerGluProGlyG1 569
2654 C- -TGTTATTCGCTGATGGCGCGCAAGCAATGAGTGGCTTTCGAGTGGCGCGCAGC 2710
569 yValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaPr 589
2711 CGGGCAGGGCGCGCGCAGCATTCCTATGTCGCGATGCGCGCAGTGGTGTGGGCACC 2770
589 oGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAla 609
2771 GCGCGCGCGCGGCAAT- -TGGCGCAGCTCCGACTTCGCGCGCAGTGGCGCGCGTGT 2824
609 rGlnGlyValProAlaAsnAlaGlnIleArgSerAspAtqValAsnProLysThrPheTy 629
2825 GAACGCCCTCCCGAACCCCGCGTGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 2884
629 rAla- -LeuSerAsnGlyThrPheTyrArgSerThrAspGlyValThrPheG1 647
2885 TGGGTGGATGTCGCGCGCGCGCAGCTGTACGAGAGCAGCAGCAGCAGCAGCAGCAGC 2944
647 nProValAlaAlaGlyLeu- - 653
2945 T- -GCGACCGGTGTACAGTGGCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 2998
654 - -ProSerSerGlyAlaValGlyValMetPheHisAlaValProGlyLysGluG1 671
2999 GCTGCGTCCGAGACCGTGGCGCGCAGCGTGGTGTATCTGGCCAGCGCGCGCGCGCAAG- 3053
671 yAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSerThrAsnGlyGlySerSerTr 691
3054 - -GGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 3082
691 pSerAlaIleThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAlaProG1 711
3083 GCAGGTGCTCTCACAGCGCGCACCAAGCAGCTGCTGGCGCATCGCGCAAGCGCTGGCGG 3142
711 ySerSerTyrProAlaValPheValValGlyThrIleGlyValThrGlyAlaTyrAr 731
3143 CGGTGGCGCGCGCGCGCTGTATCTGGCGCGCGCGTGGCGCGCGTGGATCGGTGTCG 3202
731 gSerAspAspCysGlyThrTrpValLeuIleAsnAspGlnHisGlnTyrGlyAs 751

3203 CTCGACGATGGCGCGCGCGCGCAGTGGCAGCGCATCAACGATGACGCGCCTTCGGCGG 3262
751 nTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgValTyrIleGlyThrAs 771
3263 C- -CCGTACAGCGTACCGGTGATCCGCGCATTCGCGCGGTGTGTACTTCGCCACCGG 3319
771 nGlyArgGlyIleValTyrGlyAspIleGlyAlaProSerGlySerProSe 791
3320 CGCGCGCGCATTTTTCACGGCGATCCGAGATGAGCAGTGCCTGCTCGCGCA- 3374
791 rValSerProSerAlaSerProSerLeuSerProSerProSerProSerPro- 810
3375 - -CGTCTGCTGCTTTTACCGTACCGCGCGCGCGCGCTGGCGCGTGGCGCGCGCGG 3430
811 - -SerProSerProSerProSerProSerProSerProSerProSerProSer- 827
3431 ACACCGCGGACCTTGGAGATCACCGCATGCCCGT-CCACATTTCCGCGCGCTGCTTG 3489
828 - -ProSerProSerProSerProSerProSerProSerProSerProSerProSer- 836
3490 CTGGGTGTGTGTCGCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3549
837 - -SerProSerProSerProSerProSerProSerProSerProSerProSerProSer- 850
3550 GCCGATGCGCGGTGCTGCGACGCGCATGAGCAATGCCGCTATGGGGTGGCGCGCGCGC 3609
850 oSerProSerProSerProSerProSerProSerProSerProSerProSerPro 864
3610 GCGCGCGCAATCACGGTGTGATGGCAAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCG 3661

RESULT 6

ADD42054

ID ADD42054 standard; DNA; 5698 BP.

XX AC ADD42054;

XX DT 15-JAN-2004 (first entry)

XX DE Trichoderma reesei Family 74 xyloglucanase-encoding DNA, SEQ ID NO:1 #1.

XX KW Trichoderma reesei; Family 74 xyloglucanase; beta-1,4-glycosidic linkage;
XX KW xyloglucan hydrolysis; biomass degradation; cellulose; hemicellulose;
XX KW ethanol production; detergent composition; fabric treatment;
XX OS textile treatment; ds.

XX OS Hypocrea jecorina.

XX PN WO2003089598-A2.

XX PD 30-OCT-2003.

XX PF 17-APR-2003; 2003WO-US011831.

XX PR 19-APR-2002; 2002US-0373987P.

XX PA (NOVO) NOVOZYMES BIOTECH INC.

XX PI Michael R, Zaretsky E, Haas J;

XX DR WPI; 2003-845528/78.

XX DR P-PSDB; ADD42055.

XX PT New polypeptides having Family 74 xyloglucanase activity, and encoding
XX PT nucleic acid molecules, useful for degrading cellulose- and hemicellulose
XX PT -containing biomass to ethanol or as a detergent.

XX PS Claim 1; SEQ ID NO 1; 96pp; English.

XX CC The invention relates to a Family 74 xyloglucanase (ADD42055) from the
XX CC fungus Trichoderma reesei (Hypocrea jecorina), and nucleic acids encoding
XX CC it (e.g., ADD42054). The enzyme hydrolyses the beta-1,4-glycosidic
XX CC linkages in the backbone of xyloglucan to xyloglucan oligosaccharides.

CC The invention also relates polypeptide sequences at least 70% identical
 CC acid of the enzyme, expression vectors and host cells comprising a nucleic
 CC mutant of the invention, the recombinant production of the enzyme, and
 CC mutant enzymes and the nucleic acids encoding them. The xyloglucanase of
 CC the invention can be used in the degradation of cellulose- and
 CC hemicellulose-containing biomass to produce ethanol. It can also be used
 CC in a detergent composition for treating fabric during a machine washing
 CC cycle. The present sequence represents a trichoderma reesei Family 74
 CC xyloglucanase-encoding DNA. Note: The present sequence (referred to as
 CC SEQ ID NO:1) is given as shown in the sequence listing. A comparison of
 CC this sequence with that shown in figure 1 (also referred to as SEQ ID
 CC NO:1) shows that the present sequence is twice as long as the figure 1
 CC sequence (ADD42060) and contains alternate 60 bp blocks of sense strand
 CC and complementary strand sequence.

XX Sequence 5698 BP; 1145 A; 1704 C; 1704 G; 1145 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1.04e-24 Length: 5698
 Score: 940.00 Matches: 374
 Percent Similarity: 30.11% Conservative: 136
 Best Local Similarity: 22.08% Mismatches: 315
 Query Match: 18.31% Indels: 871
 DB: 9 Gaps: 59

US-09-917-376-1 (1-957) x ADD42054 (1-5698)

Qy 25 SerPheAlaValAlaAlaLeuGlyValLeuProIleAlaIleThrAlaSerProAla 44
 Db 128 TCTCGAGTCTTCCCTTGTCTTGGGGCGGTC-----ATCCTGCC 169
 Qy 45 HisAlaAlaThrGln-----GluGlyAlaPro----- 50
 Db 170 CATGCTGCTTCTTCAGAGAGCTCAGGAACGGGAACAGACCCCGCGAGTAGGACGG 229
 Qy 51 -----ProTyrThrSerAsnValAlaIle---GlyGlyGlyPheValAspGly 67
 Db 230 GTACGCGGAATTCATGGAAGACGTCAAGCTCGCGCGCGCGCGCTTCGTCGCCGCG 289
 Qy 68 IleValPheAsn-----GluGlyAlaPro----- 75
 Db 290 ATCATCTTCCAAAGTACCTTTTCGAGTTCGAGCGCGCGCGCGCGAGCAGCGGCGCG 349
 Qy 76 -----GlyIleLeuTyrValArgThrAspIleGlyGlyMet 87
 Db 350 TAGTAGAGGTTCCCAAGACAAAGCGTAGCATATGCACGAACAGATATGGCGGCTG 409
 Qy 88 TyrArgTrp----- 90
 Db 410 TACCGCTCAAGGGTTCTGTTTCCGCATCGTATACGTGTTGCTCTATAACCGCCGAC 469
 Qy 91 -----AspAlaAlaGlyArgTrpIleProLeuLeuAspTrpVal----- 104
 Db 470 ATGCGCGAGTTCGCGACGACTCATGACCGCGCTCACGGATGGGATGCTGTAATGCC 529
 Qy 105 GlyTrp-----AsnAsnTrpGly----- 110
 Db 530 GGCTGGCACAAGCGGTGTGAGTACTCTGGCGGAGTGCCCTAACGACTATTACGG 589
 Qy 111 -----TyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysVal 127
 Db 590 CCGACCGTGTTCGGGGCATCGAGCTGTTCGCTTGATCCGACGACGATCAAAAGGTG 649
 Qy 128 TrpAlaAlaVal----- 131
 Db 650 TATGCCGAGTGAACCCCGTAGCTGCGACACCGGAACAGTAGCGGTCCTGCTAGTTTCCAC 709
 Qy 132 -----GlyMetTyrThrAsnSerTrpAsp----- 139
 Db 710 ATACGGCGTCACGCGCATGTATACGAACAGCTGTC-TGTGATGTCCTCAGATCTAGACC 768
 Qy 139 ----- 139

Db 769 TATGATTGGACGGCGTACATATGCTTGTGACACAGACACTACAGAGAGCTTAGATCTGG 828
 Qy 140 -----ProAsnAspGlyAlaIleLeuArg 147
 Db 829 ATACTAACCTGCCTGACATTGGCCATATATAGGGATCCGAGTAATGGAGCCATCAITCGC 888
 Qy 148 SerSerAspGln----- 151
 Db 889 TCGTCAGACCGCGACTGTAAACCGGTATATATCCCTAGGCTCAATTACCTCGGTAGTAGCG 948
 Qy 152 -----GlyAlaAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsn 167
 Db 949 AGCAGTCTGGCGGCGCAACGTTGCTTCCACAACTTGCCTTCAAGTTCGGGGTAAAC 1008
 Qy 168 MetProGlyArg----- 171
 Db 1009 ATGCCAGACCGCGCGTTGCACAGGAAGTGGTTGAACGGGAAGTTTCAGCCCCCATTTG 1068
 Qy 172 -----GlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeu 187
 Db 1069 TACGTCCTCGCGGAGCGGAGAGCGTCTGCTGTGATCCGCGCAACTCCAAACATCATC 1128
 Qy 188 TyrPheGlyAlaPro----- 192
 Db 1129 TACTTTGGTGTCTCTCGGCTCTCCACAGACCGACAGCTAGGCCGTTGAGTTGTAGTAG 1188
 Qy 193 -----SerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrp 207
 Db 1189 ATGAACACACAGCGCTCAGGAACCGCTCTGGAAAGTCTACGAGCGCGCGGTGACCTTT 1248
 Qy 208 SerGlnMet----- 210
 Db 1249 TCCAGGTCCTCGCGAGTCTTTTGCAGGAGACCTTCAGATGCTGTCGCGCGCACTGGAAA 1308
 Qy 211 -----ThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThr 227
 Db 1309 AGGTTCCAGAGCTGTTTCAACGCAACTGGGAGCTATATCCAGACCCGAGTGTATCCAAC 1368
 Qy 228 GlyTyrGln----- 230
 Db 1369 GGCTAACACAGCAGCAAGTCCGTTGACCTGCATGTAGGTCCTGGGCTCACTAAAGTTG 1428
 Qy 231 -----SerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerSer 247
 Db 1429 CCGATGTTGTCGACACAGCAAGAGTCTATGTGGTTCAGTTCGACTCAACACGACGACG 1488
 Qy 248 LeuGlyGlnAla----- 251
 Db 1489 ACCGGGGAGCGCTTGTCTGTTCTGAGTACACCCCAATGCAAGCTGAGTTGGTCTCGTGC 1548
 Qy 252 -----SerLysThrIlePheValGlyValAlaAlaAspProAsnAsnProValPhe 267
 Db 1549 TGGCCCCCTCGAGCGTCTGTATCTTTGTTGGCAGCGGTGATAACATCACTGCTTCAGTC 1608
 Qy 268 Trp----- 268
 Db 1609 TATGTGACAGCTGCAGAGCATAGAAACACCGTCCCGACTATTGTAGTGACGAAGTCAG 1668
 Qy 269 -----SerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPhe 286
 Db 1669 ATACACTCGTGAATGCCGCTCCACGTTGGAGTGTGTATACCGGGGCGAGCCAGGAAATAC 1728
 Qy 287 IleProHis----- 289
 Db 1729 TTTCTCACAAGTTTACGGCCGAGGTGCACCTCAGACATGCGCCCGTCCGTTCCCTTTATG 1788
 Qy 290 LysGlyValPhe-----AspProValAsnHisValLeuTyrIleAlaThrSerAsn 306
 Db 1789 AAGAGGTGTTCCGGAACACTGCAGCCGAGGAAGGCGCTTGTATCTGACCTATTCCGAT 1848
 Qy 307 ThrGlyGlyPro----- 310
 Db 1849 GGCACAGGGCGCGCTTTTGACGTGGTCTCTTCCGGAACATAGACTGGATAAGGCTA 1908

Qy 311 -----TyrAspGlySerSerGlyAspValTrpIysPheSerValThrSerGly 326
Db 1909 CGGTGTCGCGGCTATCATGGCACACTTGGCTCAGTGTGGAGGTACGACATTGCGAGGGGA 1968
Qy 327 ThrTrpThrArgIleSerProValProSerThrAspThr----- 339
Db 1969 ACTTGGAAAGACATACTTACCGTGTGAACCGAGTGCACACTCCCATGCTGTAACTGCCCT 2028
Qy 340 -----AlaAsnAspTyrPheGlyTyrSerGlyLeu 349
Db 2029 TGAACCTTTCTGATCACCCTGTCTCTGGATCAGATCTATCTTTGGCTTTGGCGGCTT 2088
Qy 350 ThrIleAsp----- 352
Db 2089 GGCCTCGATTGTAGTGGGACAGAGACCTAGTCTAGATATGAACCGAAACCGCGGAA 2148
Qy 353 -----ArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrpPro 369
Db 2149 CCGAGACTAAACCAAAAGCCAGGAACCTTGTGTGCTTCTTTGAACTCTTGTGGCCA 2208
Qy 370 AspThr----- 371
Db 2209 GATGCTCAGCTGTTTGGTCTCTTGGGAACAACAGAAACTTGAACAACCGCGT 2268
Qy 372 -----IleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIleTrpAspTrp 389
Db 2269 CTACGAGTCGACTTCTCGTGCACCGACTCTGGGACACATGGAGCCGATCTGGCGGTG 2328
Qy 390 ThrSerTyrProAsnArgSer-----LeuArgTyrValLeuAspIleSerAlaGluPro--- 407
Db 2329 GCGAGCTATCCGAAAGCCAGCTGGCTGAGACC-CTGTTGTACCTCGGGCTAGAACCCGAC 2387
Qy 407 ----- 407
Db 2388 CCGCTCGATAGGCACTGAGACCTATTACTACAGCATCTCAGTAGTCACCTCTTAAACGATC 2447
Qy 408 -----TrpLeuThrPhe-----GlyValGlnProAsn----- 416
Db 2448 CGATGCGAATGGCTGACTCTGGATAATGATGCTCCTAGAGTCACCTCAGTGAGATGTGTAG 2507
Qy 417 -----ProProValPro----- 420
Db 2508 GCTACGCTTACCGACTGACTCTGCTTCAAGACTCCCAAAAGCACCGTGGATCAAGAACAAC 2567
Qy 420 ----- 420
Db 2568 TTTATCGATGTGATGACTGAGACGAAGTCTGAGGGTTTCGTGGCACTAGTCTTCTTGTG 2627
Qy 421 -----SerPro-----LysLeuGlyTrpMet 427
Db 2628 AATAGCTACACTCGAGCGAGTCACCGTCCGATGGTCTCATCAAGCGCTCGGCTGGATG 2687
Qy 428 AspGluAlaMetAla----- 432
Db 2688 ATTGAGTCTCTCGGCTCGCTCAGTGGCAGGCTACACAGTAGTTCGGGAGCGGACCTAC 2747
Qy 433 -----IleAspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAla 447
Db 2748 TAACTCAGAGAGCAGATTGACCAACTGGGACACCGACCAACCACTGGCTCTACGCGCAGGAATG 2807
Qy 448 ThrLeuTyrAlaThr----- 452
Db 2808 ACAATCTTTGGCGCTAACTGGGTTGGCTGCTGTTGGTGACCGAGATGCCGTGGCCTTAC 2867
Qy 453 -----AsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAla 467
Db 2868 TGTTAGAAACCGCGCCACGATCTCACCAACTGGGACACCGCCACAATGTGTCAATCAA 2927
Qy 468 ProMetValLysGly----- 472
Db 2928 TCACTGGCAGACCGCGTCTAGAGTGGTGTGACCCCTGTGCGGGTGTATACACAGTTAGTT 2987

Qy 473 -----LeuGluGluThrAlaValAsnAspLeuIleSerProProSerGly 487
Db 2988 AGTGACCGTCTGCGCATCGAGGAATTCTCCGTCCAGGACCTGGCTGCACACCGCGGA 3047
Qy 488 AlaProLeuIleSer----- 492
Db 3048 AGCGAGCTATTGGCTAGTCTCTTAAGAGGCGAGTCTCTGGACCGGAGAGCTGGGCGGCT 3107
Qy 493 -----AlaLeuGlyAspLeuGlyPheThrHisAlaAspValThrAla 507
Db 3108 TCGCTCGATAAACC CGCAGTCGGAGACGACAACGGCTTACCTTTGCCACGAGAAACGAC 3167
Qy 508 Val----- 508
Db 3168 CTGGGACATCGCGCGCTCAGCCTCTGCTGTTCGGAAGTGGAAACGGTCTGCTTTGTG 3227
Qy 509 -----ProSerThrIlePheThrSerProValPheThrThrGlyThrSerVal 524
Db 3228 GAGCCCTGTAGCGCGACAGCTCTGGCAACGCCCATATGGGCCACCTTCGACGAGCGTC 3287
Qy 525 AspTyrAla----- 527
Db 3288 GACTACCGCGGAGCGTCTGCCAGCCGTTGCGGTTGTAACCGGTGGAGCTGCTCGCAG 3347
Qy 528 -----GluLeuAsnProSerIleIleValArgAlaGlySer----- 539
Db 3348 CTGATCGCGGCTACTCGTCAAGAGCGTCTCGCGTCGGCAACACCGCCGCGCACGAA 3407
Qy 540 -----PheAspProSerSerGlnProAsnAspArgHisValAla 552
Db 3408 CAGGTGGCCATTTGAGCCAGTCTTCGACGAGCGCGACCGTTGTGGCGCGCGTGTGCGTT 3467
Qy 553 Phe-----SerThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGly 569
Db 3468 GTCCACCGGTAGACTCGACGCGCGCGCGAGTGCATCGACTACGCGGCCACACG 3527
Qy 570 ValThrThrGlyGly----- 574
Db 3528 TCCATGAACGCGCGCAGGCTCGCGCGCTGCACCTCGTAGCTGATCGCGCGGTGTGC 3587
Qy 575 -----ThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaPro 589
Db 3588 AGGTACTTGGCGCACCGTGGCTATTTCGCGCGACGCGCACCATCTCTGGTGCACC 3647
Qy 590 GlyAspProGly----- 593
Db 3648 GCCTCTCGCGCGCTGCCACCGGATAGCCGCTGCGCTGCTGTAGGAGACCGAGTGG 3707
Qy 594 -----GlnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGlnGly 611
Db 3708 CGGAGCAGCGCGCTGCAGCGCTCG- -CAGTTCCAGGCGCAGCTTTGCTCTCGTCTCGAGC 3764
Qy 612 ValProAlaAlaAlaGln----- 617
Db 3765 CTGCGCGCGCGCGCAGCTCGCGAGCGTCAAGTTCGCTCGAAACGGAGGCGAGCTCG 3824
Qy 618 -----IleArgSerAspArgValAsnProLysThrPheTyrAlaLeu 631
Db 3825 GACGGCGCGCGCGCTCATCGCTCGGACAAGAAGAACCAACAGCGCTCTTACGCCGCG 3884
Qy 632 SerAsnGlyThrPhe----- 636
Db 3885 TCCGATCGACCTTTTAGTAGCGGAGCTGTTCTTCTGTTGTCGACAGAGATGCGGCG 3944
Qy 637 -----TyrArgSerThrAspGlyValThrPheGlnProValAlaAla 651
Db 3945 AGGCTAGCTGGAATAACGTCAGCAAGGACACCGCGCAGCAGCTTC- -ACGCGC 3995
Qy 652 GlyLeuProSerSerGlyAlaValGlyValMet----- 662
Db 3996 GGG- -CCCAAGCTGGGCGCAGCGAGGATGAGTCTTCTGTCGCGCGTGTGCGAGTGC 4052
Qy 663 -----PheHisAlaValProGlyLys 669

352 GGCTTACGACATATCTGGAGCGCGCTCTCCAG-----CAGCAGGA 396
466 IleAlaProMetValLysGlyLeuGluGluThrAlaValAsnAspLeuIleSerProPro 485
397 ACCACCCCTGAAGTC-----ATTGAGATTGTCTATCTACGATCTG-----CGGGA 441
486 SerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspVal 505
442 CGGACTGCGCGCGCTCGCTCTCAACGCGGAATCGCCGCT---ACGGCAGCAGAGTTG 498
506 ThrAlaValProSerThrIlePheThrSerProValPheThrGlyThrSerValAsp 525
499 CAGACCTAT---GAAACGAGTATCATCGATCCGATTCGCGATATCTCGACCAATCCGAAG 555
526 TyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSerSerGln 545
556 TACTCCAGCTCGGATCGTGACGATCATI-----GAGCCGAGCTCGCTG 600
546 ProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySer 555
601 CCAACGCG-GGTCA-----CAATAT 620
566 GluProGlyGlyValThrThrGlyGlyThrValAla-----AlaSerAlaAspGly 582
621 GAGATTCAAGCTGTGCAACGCGGTGCGGTATT-ACGAGCAAGGCATCGAGTACGCGC 679
583 SerArg-----PheValTrpAlaProGlyAspPro--- 592
680 TCAGCAATTCACGCCATTCGNAACGTGTACATCTACATGGAGCGCGCCACTCCGGCT 739
593 -----GlyGlnProValVal-----TyrAlaValGlyPheGlyAsnSerTrp 606
740 GGCTTGGTGGCCCAATAATACCGAGCGGATACGTACAGGAAGTCCAGAAAGTCTCAACG 799
607 AlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsn----- 624
800 CGAGCATCGGGGTCAACGGCATCGAGCGCTCTGTCCACCAACACGCGCAATTACACGCCGT 859
625 -----ProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyr 637
860 TGAAGGAGCGTTCATGACCGCCACCCAGCGAGCTCGCGGACACCGCGTGGAGTTCGCGGA 919
638 ArgSerThrAspGlyGlyValThrPheGlnPro----- 648
920 ATTTCTACCACTGGAAPCTCTGACATC-GACGAAGCCGACTACGCGGTGACTTGTACTCG 978
649 -----ValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaVal 666
979 CGGCTCGTCCCGCTGGCTTTCAAGCAGC-----ATCGGCATGCTCATCGACACCTTA 1032
667 ProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSerThrAsn 686
1033 -----CGCAC 1038
687 GlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnVal----- 703
1039 GGTGGGGTGGTCCGAACGAACCAACAGCCGCGACACCGCGACCGCGATGTCAACACCTTC 1098
704 -----Gly 704
1099 GTCAACAGCTCGAAGATTGACCTTCGCGACGACCGCGCTGTGGTGCACACAGACCGT 1158
705 PheGlyLysSerAlaProGlySerSerTyrProAlaValPhe----- 718
1159 GCGGGCTTCGCGCAGCGCGGCAAGCGCGAGACTTCCCGAAGCGCGACCTCGAC 1218
719 -----ValValGlyThrIleGlyGlyValThrGlyAla----- 729
1219 GCGTATGCTCGATCAAGCGCGGGTGTAGTCGACGCGCACAGCGGTGCGAGCGATCCG 1278
730 -----TyrArgSerAspAsp---CysGlyThrTrpValLeuIleAsnAspAsp 745
1279 ACAACTGGCAAGAGTCGAGCCCATGTGCGACCGCGAGCTACACGAGC----- 1326

Qy 746 GlnHisGlnTyrGlyAsnTrpGlyGlnAla-----IleThrGlyAspHis 760
Db 1327 -----TCGTACGGGGTACTGACCAACGCGTACTCCGAACCTCCCGATCGCGCCAGTGG 1380
Qy 761 AlaAsnLeuArgArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIle 780
Db 1381 TTCCCGCGCGAGTTTGGACCGTTCGCGAACGACGCGCGTCCGACGCTCGGAC 1440
Qy 781 GlyGly-----AlaProSerGlySerProSerProSerValSer 793
Db 1441 AGCTCGAGCGCGCTCCCGCCGAGTCCGTCGGTTCGCGAGTCCGAGCCCGAGT 1500
Qy 794 ProSerAlaSerProSerLeuSerProSerProSerProSerProSerProSerProSer 813
Db 1501 CCGAGCCGCGAGCTCGCATCGCGTTCGCGAGTTCGAGCTCGAGCCGCTCTCCGTCG 1560
Qy 814 Pro-----SerProSerSerProSerProSerProSerProSer 825
Db 1561 CCGAGCCGAGTCCGAGCCCGAGTTCGCGCGTCCGCTCTCCGAGCTCGAGCCCGTCT 1620
Qy 826 ProSerProSerProSerProSerArgSerProSerProSerProSerAlaSerProSer 845
Db 1621 CCGTCGCGAGCCCGAGTTCGAGCCCGAGTTCGCGTTCGCGCTCTCCGAGCTCGAGC 1680
Qy 846 ProSerSerProSerProSerProSerProSerProSerProSerProSerProSerProSer 865
Db 1681 CCGTCTCCGTCGCGAGCCCGAGTCCGAGCCCGAGTTCGCGCTCGCGCTCTCCGAGC 1740
Qy 866 SerSerProValSerGlyGlyValIleValGlnTyrLysAsnAsnAspSerAlaProGly 885
Db 1741 TCGTCGCGGTTCGCGTGGGTGAAGTGCAGTACAGAAACAATGATTCGCGCGCGGT 1800
Qy 886 AspAsnGlnIleLysProGlyLeuGlnValValAsnThrGlySerSerSerValAspLeu 905
Db 1801 GATAACCAAGATCAACCGGGTCTCAAGTGTGTGATATACCGGTCTCGTGGTGGATTG 1860
Qy 906 SerThrValThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyr 925
Db 1861 TCGACGGTTCGCGTACTGGTTTCAACCGGATGGTGGTCTCGACACTGGGTGTAC 1920
Qy 926 AsnCysAspTrpAlaAlaIleGlyCysGlyAsnIleArgAlaSerPheGlySerValAsn 945
Db 1921 AACTGTGACTGGCGCGCATGGGGTGTGGGAATATCCGCGCTCGTTCGGCTCGGTGAAC 1980
Qy 946 ProAlaThrProThrAlaAspThrTyrLeuGln 956
Db 1981 CCGCGAGCGCGACGCGGACACCTTACCTGCAG 2013
RESULT 8
ABZ77633
ID ABZ77633 standard; DNA; 2289 BP.
XX ABZ77633;
XX ABZ77633;
DT 03-JUN-2003 (first entry)
XX
DE Nucleotide sequence of the Mana polypeptide.
XX
KW Mana; mannanase A; glycoside hydrolase; enzyme; hemicellulose; sugar;
XX food; feed; paper pulp; biofuel; mannanase; gene; ss.
XX Acidothermus cellulolyticus.
XX
FH Key Location/Qualifiers
FT CDS 1..2289
FT /*tag= a
FT /product= "Mana"
XX
XX WO2003012110-A1.
XX
PD 13-FEB-2003.
XX

PF 28-JUL-2001; 2001WO-US023819.
 XX PR 28-JUL-2001; 2001WO-US023819.
 XX PA (MIDE) MIDWEST RES INST.
 XX FI Ding S, Adney WS, Vinzant TB, Himmel ME;
 XX DR WPI; 2003-248182/24.
 XX DR P-PSDB; ABP73022.
 XX PT Novel thermal tolerant mannanase A polypeptide derived from Acidothermus
 XX PT cellulolyticus, useful for reducing hemicellulose in a starting material,
 XX PT for processing of food, and as bulking agents in food stuffs.
 XX PS Example 1; Page 23; 46pp; English.
 XX CC The present sequence encodes ManA, a thermostable mannanase A polypeptide
 XX CC derived from Acidothermus cellulolyticus. ManA is a member of the
 XX CC glycoside hydrolase family of enzymes. ManA is useful for reducing
 XX CC hemicellulose in a starting material to simpler carbohydrate units, and
 XX CC ultimately to sugars which are useful in the food, feed, paper pulp, and
 XX CC biofuels industries. It is useful for the processing of food and in food
 XX CC stuffs as bulking agents, and for the degradation of mannan. ManA is
 XX CC also useful to raise polyclonal and monoclonal antibodies that are useful
 XX CC in purifying ManA, or detecting ManA polypeptide expression, and as well
 XX CC as reagent tools for characterizing the molecular actions of ManA
 XX CC polypeptides
 XX SQ Sequence 2289 BP; 463 A; 700 C; 688 G; 438 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1,17e-16 Length: 2289
 Score: 702.00 Matches: 214
 Percent Similarity: 43.70% Conservative: 53
 Best Local Similarity: 35.02% Mismatches: 148
 Query Match: 13.67% Indels: 196
 DB: 7 Gaps: 27

US-09-917-376-1 (1-957) x ABZ77633 (1-2289)

Qy 472 GlyLeuGluGluThrAlaValAsnAspLeuSerProSerGlyAlaProLeuLeu 491
 Db 4 GGTCTAGTCGTCCTCGCGAGCATTTGTCGACCGCGCGGCACT---GCGGTT 60
 Qy 492 SerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAlaValProSerThr 511
 Db 61 GCTGCGCGCGAGCGTCTCGGCTCAATCACCATGCCGTCAGCCAGCGCGCGCGGGA 120
 Qy 512 IlePheThrSer-----ProValPheThrGly 521
 Db 121 TTGCTACCGCATCCGCGGTCAGTTCGTTCTGAAACGGCTTCCTATCCTTACGGGGGA 180
 Qy 522 ThrSerValAspTyrAlaGlu----- 528
 Db 181 ACGAACACTATTACCTCAGCTATCAGTCGACGCGCGCATGATGATGTTGGCCAG 240
 Qy 529 -----LeuAsnProSerIlele-----ValArgAlaGlySerPhe 540
 Db 241 GCTCAAGCGATGAATCTTTCTGTATCCGACCTGGGGTTTTCATCGACATCGGCTCTCT 300
 Qy 541 AspProSerSerGlnProAsnAspArgHisAlaAlaPheSerThrAspGlyGlyLysAsn 560
 Db 301 GAC---GGCTCGTGCGCC-----ACAAATCGATGCAACAAGAAC 336
 Qy 561 -----TriPheGlnGlySerGluProGlyValThrThrGly----- 573
 Db 337 GGCTTCTACTTTCAGTACTGGGACCG-----TCGACCGCGCTCCGGGTACAC 387
 Qy 574 -----GlyThrValAlaAlaSerAlaAspGly 582
 Db 388 GACGGGCGGACCGGCTTGCAAGGCTTGAAGCTTGAAGCTGCGAGCGCGCGCGCGGCG 447

Qy 583 SerArgPheValTrpAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPhe 602
 Db 448 CTTCCGGTGATT-----GTCGTCTCTC 468
 Qy 603 GlyAsnSerTrpAlaAlaSer-----GlnGlyValPro 613
 Db 469 ACCAACGACTGGAAAGAAATTTGGGGGAATGATCAATACGACAACTGGTACGGCTTCTCT 528
 Qy 614 AlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyrAlaLeuSerAsn 633
 Db 529 TACCACGACACTTCTACACCGAC-----CCCGGACCCAGCAGCGGTACAGAT 579
 Qy 634 GlyThrPheTyrArg-----SerThrAspGlyGlyValThrPhe----- 646
 Db 580 TGGGTCAATCATCTACTGAAACCGGGTCAACAGCATTTACCGGCGTCACTGACGTACAAGAAGCAT 639
 Qy 647 -----GlnProValAlaAlaGly-----Leu 653
 Db 640 CCAACGATCTTGTCTGGGAACCTTGGCAATGAGCGCGCTGCTAGGAAGCGGCACATT 699
 Qy 654 ProSerSerGlyAla-----ValGlyValMetPheHis 664
 Db 700 CCAACCTCGGCGACGTGACTCAGCGGACCAATGTCAACTGGTGCATCAAAATGTCGGCG 759
 Qy 665 AlaValProGlyLysGluGlyAspLeuTrpLeuAla-----AlaSerSerGlyLeuTyrHis 683
 Db 760 TACGTCAAAAGCATAGACCCCTAACCATATGCTCTCGGTCCGCGACGAAGGGTTCTACATT 819
 Qy 684 SerThrAsnGlyGlySerSerTrp-----SerAlaIleThrGlyValSerSer----- 699
 Db 820 GGGTCAACGAGGAGGCGGCTGCCATACACGACCGCTCGACGGCTGCACAAACAT 879
 Qy 700 -----AlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrPro 715
 Db 880 GCTCTTCTCGGTGTCAGAAACATTGACTTTGCGCACGTAT-----CACCTGTACCCG 930
 Qy 716 AlaValPheValGlyThrIleGlyGlyValThrGlyAlaTyrArgSerAspAspCys 735
 Db 931 AATTACTGG-----GGCCAGAACGCGACTGG 957
 Qy 736 GlyThrThrTrpVal-----LeuIleAsn----- 743
 Db 958 GGAACGCAATGATCAAGGATCATATTGCGAATCGCGAGCGATCGGCAAGCGGACCAT 1017
 Qy 744 -----AspAspGlnHisGlnTyrGlyAsnTrpGlyGln 754
 Db 1018 CTCGAAGAATTCGGCTGGCAGACACCGAGCGCATTCCTCTATCAGACGTGGACCCAG 1077
 Qy 755 AlaIleThrGlyAspHisAlaAsnLeuArgValTyrIleGlyThrAsnGlyArg--- 773
 Db 1078 ACTGTGCGT-----ACGAACGCTGAAGCA 1101
 Qy 774 GlyIleValTyrGlyAspIleGlyAlaProSerGlySerProSerProSerVal--- 792
 Db 1102 GGCTGGAACCTCTCGATGCTCGCTGGGAATGTCACAGCGCCAGCATATCCGAACATGAC 1161
 Qy 793 -----SerProSerAlaSerProSerLeuSerProSerProSer 805
 Db 1162 GGCTTCAACGCTACTACTACCAAGTTTCAACAGCAGCCGCTCGCAGCGAGCGCTCGCA 1221
 Qy 806 ProSerSerProSerProSerProSerProSerProSerProSerProSer 825
 Db 1222 ATCAGTACCAGGACATCGCTCCGCGCGTCCGCGAGTTCAGTCCATCTCTCGTCCGCGCT 1281
 Qy 826 ProSerProSerProSerProSerProSerProSerProSerProSerAlaSerProSer 845
 Db 1282 CCGTCCGCGCTCTCGTGGCGTCTCCGTCGCGCTCCG----- 1320
 Qy 846 ProSerSerProSerProSerProSerProSerProSerProSerProSerProSer 865
 Db 1321 -----TCGCGCTCTTCGTCGCGAGCCGCTCTCCGCTCG 1353
 Qy 866 SerSerProValSerGlyGlyValValGlyValGlnTyrLysAsnAspSerAlaProGly 885

Db 1354 TCGTCCCGGTCGGTGGGTGAAGTGCAGTCAAGAACAAATGATTCGGCGCCGGGT 1413

Qy 886 AspAsnGlnIleIysProGlyLeuGlnValValAsnThrGlySerSerValAspLeu 905

Db 1414 GATACACAGATCAACCGGGTCTCGAGTGGTGAATACGGGTCGTCGGTGGATTG 1473

Qy 906 SerThrValThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyr 925

Db 1474 TCGACGGTACGGTGGTACTGTTTCCACCGGGATGGTGGTGGTGGTGGTGGTGGTAC 1533

Qy 926 AsnCysAspTrpAlaAlaIleGlyCysGlyAsnIleArgAlaSerPheGlySerValAsn 945

Db 1534 AACTGTGACTGGCGCGGATGGGGTGGGAATATCCGGCCTCGTTCGGCTCGGTGAAC 1593

Qy 946 ProAlaThrProThrAlaAspThrTyrLeuGln 956

Db 1594 CCGCGACCGCGGACGGGACACTTACCTGCAG 1626

RESULT 9

ID ABZ76162 standard; DNA; 3365 BP.

XX AC ABZ76162;

XX

DT 29-MAY-2003 (first entry)

DE A. cellulolyticus Gux1 protein encoding DNA.

XX Gux1; thermal tolerant; exoglucanase; glycoside hydrolase; cellulose;

KW biofuel; detergent; pulp; paper processing; feed processing; textile;

KW cellulase; gene; ds.

XX Acidothermus cellulolyticus.

XX

FT Key Location/Qualifiers

FT CDS 1..3365

FT FT /*tag= a

FT FT /product= "Gux1"

FT FT /transl_except= (pos: 682..683, aa: Pro)

FT FT /notes= "this codon has an apparent one nucleotide

FT FT basepair deletion which alters the reading frame"

XX

PN WO2003012095-A1.

XX

PD 13-FEB-2003.

XX

PF 28-JUL-2001; 2001WO-US023820.

XX

PR 28-JUL-2001; 2001WO-US023820.

XX

PA (MIDE) MIDWEST RES INST.

XX

PI Adney WS, Ding S, Vinzant TB, Himmel ME, Decker SR;

PI Lantz McCarter S;

XX

DR WPI; 2003-300494/29.

DR P-PSDB; ABP71656.

XX

XX New thermal tolerant Gux1 peptide having specified amino acid sequence,

PT useful in the degradation of cellulose to biofuels.

PT

XX Disclosure; Page 22-23; 44pp; English.

XX

CC The invention relates to a thermal tolerant Gux1 peptide from A.

CC cellulolyticus. The Gux1 exoglucanase is a member of the glycoside

CC hydrolase family and comprises a catalytic domain GH48, carbohydrate

CC binding domain type III, and a carbohydrate binding domain type II. The

CC polypeptide is useful in the degradation of cellulose into biofuel, or

CC for conversion of biomass to biofuel additives. It is used in detergents,

CC pulp and paper processing, food and feed processing, and in textile

CC processing. It can also be used alone or in combination with other

CC cellulase or glycoside hydrolases. The novel polypeptide generates

CC alternative cellulase enzymes capable of assisting in the commercial-

CC scale processing of cellulose to sugar for use in biofuel production. The

CC present sequence represents a A. cellulolyticus Gux1 cellulase encoding

CC DNA

° SQ Sequence 3365 BP; 669 A; 1117 C; 1041 G; 538 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 6.68e-07 Length: 3365

Score: 429.00 Matches: 79

Percent Similarity: 97.62% Conservative: 3

Best Local Similarity: 94.05% Mismatches: 2

Query Match: 8.35% Indels: 0

DB: 7 Gaps: 0

US-09-917-376-1 (1-957) x ABZ76162 (1-3365)

Qy 873 ValIysValGlnTyrIlysAsnAsnAspSerAlaProGlyAspAsnGlnIleIysProGly 892

Db 112 CTCAACGCGCAGTATAAGAACATGATTCGGCGCGGAGTGCACACCAACGATCAACCGGGT 171

Qy 893 LeuGlnValValAsnThrGlySerSerValAspLeuSerThrValThrValArgTyr 912

Db 172 CTCAGTTGGTGAATACCGGGTCGTCGGTGGATTGTGACCGGTGCGGTAC 231

Qy 913 TrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAlaIle 932

Db 232 TGGTTCCACCGGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 291

Qy 933 GlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAlaAsp 952

Db 292 GGGTGTGGGAATATCCGGCCTCGTTCGGTGGTGAACCGCGGACCGCGCGCGGAC 351

Qy 953 ThrTyrLeuGln 956

Db 352 ACCTACCTGCAG 363

RESULT 10

AAK55661

ID AAK55661 standard; DNA; 11706 BP.

XX

AC AAK55661;

XX

DT 30-JUL-1999 (first entry)

XX

DE DNA sequence encoding truncated cellulases Cel B4/5 and Cel B5.

XX

KW Cellulase; proteinase; truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;

KW Cel 1/2/3; Cel 6; Cel E3/B5; laundry detergent; stain-removing;

KW cotton-containing fabric; stonewashing; ss.

XX

OS Unidentified.

XX

FN EP921188-A2.

XX

PD 09-JUN-1999.

XX

PF 15-SEP-1998; 98EP-00810919.

XX

PR 19-SEP-1997; 97US-00932571.

XX

XX (CLRN) CLARIANT FINANCE BVI LTD.

XX

PI Anderson P, Bergquist PL, Daniels RM, Farrington GK, Gibbs MD;

PI Morgan H, Williams DP;

XX

DR WPI; 1999-315403/27.

DR P-PSDB; AAY13492.

XX

XX New truncated cellulase proteins, useful in detergents and for producing

PT 'stonewashed' denim.

XX

PS Disclosure; Page 20-23; 65pp; English.

XX The invention relates to a recombinant cellulase active protein free of
 CC proteinases of native thermophilic and alkaliphilic origin, comprising
 CC the truncated sequences Cel B5, Cel B4/5, Cel E1, Cel E1/2, Cel E1/2/3,
 CC Cel 6 or Cel E3/B5, or a stability region from one of the defined full-
 CC length sequences, or functional equivalents. Cel B5 extends from amino
 CC acid A1011 to F1424 or K1425 or N1426, and Cel B4/5 extends from amino
 CC acid K635 to N1426 in the sequence shown in AAY13492; Cel E1 extends from
 CC amino acid Y39 to D481, Cel E1/2 extends from Y39 to G635, Cel E1/2/3
 CC extends from Y39 to G812, Cel E6 extends from amino acid V1233 to K1751
 CC and the stability region extends from amino acid E482 to G635 in the
 CC sequence shown in AAY13493; Cel E3/B5 is shown in AAY13494. The new
 CC enzymes are useful in laundry detergent compositions to prevent or remove
 CC staining, backstaining or graying, for use on cellulosic materials
 CC including cotton-containing fabrics. They are especially useful for
 CC preventing redeposition of colorant during stonewashing, and for
 CC processing of textiles where cellulose breakdown is required. The new
 CC truncated enzymes show reduced redeposition of dye compared to using non-
 CC truncated cellulase compositions
 XX
 SQ Sequence 11706 BP; 3828 A; 1994 C; 2994 G; 2890 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 9.59e-06 Length: 11706
 Score: 409.00 Matches: 263
 Percent Similarity: 32.37% Conservative: 163
 Best Local Similarity: 19.98% Mismatches: 386
 Query Match: 7.96% Indels: 504
 DB: 2 Gaps: 58

US-09-917-376-1 (1-957) x AAX55661 (1-11706)

QY 14 ArgArgLeuValSerLeuLeuAlaAlaThrAlaSerPheAlaValAlaAlaLeuGly 33
 DB 735 AAGAGGGTAATTCATCTTT---TCCTTATGTTT-----TAAACACGCTGTAGGT 791
 QY 34 ValLeuProIleAlaIleThrAlaSerProAlaHieAlaAlaThrGlnProTyrThr 53
 DB 792 ACTTTG-----ATATTTTCATCAGGAGCAAGACGACGCA-----TATACT 833
 QY 54 TrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsnGluGly 73
 DB 834 -----GTTGATTTTGAAGGT 848
 QY 74 AlaProGlyIleLeuTyr-----ValArgThrAspIleGlyGly 86
 DB 849 GCTGATACTTTTATCTTACTTTGCTTATGGAATTCGAGCATACAGTTGACATGGGCAAT 908
 QY 87 MetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrp 106
 DB 908 ----- 908
 QY 107 AsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLys 126
 DB 909 -----GCATATAATGCTAAAGTAGTGTCTCAGGCTGTCA-----AATAGAAGTTCA 953
 QY 127 ValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu 146
 DB 954 ATATGGGATGGAGTTGCAGTT-----GACGTTAAA 983
 QY 147 ArgSerSerAspGlnGlyAlaThrTrpGlnIleThr----- 158
 DB 984 AACATTATGAACAAATGAACCATGCGGTATGTTTCAGCGGTATGTAAACATAGCTACCAG 1043
 QY 159 ---ProLeuProPheLysLeu-----GlyGlyAsnMetProGlyArg 171
 DB 1044 AAGCCGGTTCATTTGTTATCTCAGCGGTTTACGACGATCGAAGTGGGTTAAGACTACT 1103
 QY 172 GlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAla 191
 DB 1104 CTCATAGTGGAGTTGTGGCTATT---CCAAAT----- 1133
 QY 192 ProSerGlyLysGlyLeuTyrArgSerThrAspSerGlyAlaThrTrpSerGlnMetThr 211

DB 1134 -----TATGGAGAAATTT----- 1148
 QY 212 AsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGlnSer 231
 DB 1149 -----GTTGGTAATGGACTCAATATTAGCAATGTCCAGGAATTTGTTAATT 1196
 QY 232 AspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerLeuGlyGlnAla 251
 DB 1197 GTAATACACACAAATGTAGAAAGCGAGTAGATATAATGTTGAC----- 1241
 QY 252 SerLysThrIlePheValGlyValAlaAspProAsn-----AsnProValPhe 267
 DB 1242 -----TATATCCAAATAATGATGATATAATAGTTACTTATCAAAATCGAGTGACA 1289
 QY 268 TrpSerArg-----AspGlyGlyAlaThr-----TrpGlnAla-----Val 279
 DB 1290 TTTTCAAGTGGATTTGAAGTGGCACTACCGAGGTTTGGCAGGCAAGGGAACCGGTGTT 1349
 QY 280 ProGlyAlaProThrGlyPheIleProHieLysGlyValPheAspProValAsnHisVal 299
 DB 1350 ACAGTAAACCCAGATAGCGTTGTGSCATATAGTGGCAAGTATAGT----- 1394
 QY 300 LeuTyrIleAla-----ThrSerAsnThrGlyGly-----ProTyrAspGly----- 313
 DB 1395 TTGTACGTCAGTGGAGAAACGCTCAAAATGGCATGGTGCACAGATTCGCGTAGATACAATT 1454
 QY 314 ---SerSerGlyAspValTrpLysPheSerVal-----ThrSerGlyThrTrp 328
 DB 1455 TTGAAACACGGGTAAAGTGTATAAATAAGTTTGGGTTTATCAGAACACGTTGGTCAACT 1514
 QY 329 ThrArgIleSerProValProSerThrAspThrAlaAsnAsp---TyrPheGlyTyrSer 347
 DB 1515 CAAAAATGTCATTAACTATCCAAAGAGAGATTGTCTACAGATCCTTCAACAAGCTATGAA 1574
 QY 348 GlyLeuThrIleAspArgGlnHieProAsnThrIleMetValAlaThrGlnIleSerTrp 367
 DB 1575 AATCTGATATATAACAGGGATGTACCGAGTAATACGTGGGTTGAGCTGAGTGGAGCTAC 1634
 QY 368 TrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIleTrp 387
 DB 1635 -----TCATTCCTGCTGCTGTGTACAGTTAGCAGGTTGTTG 1670
 QY 388 AspTrpThrSerTyrProAsnArgSerLeuArgTyr-----ValLeu 401
 DB 1671 CTTTATGTTGAGGCACAAAATGCAAAATTTGCTTCTCGGTTGATGATTTAAGATTAT 1730
 QY 402 AspIleSer-----AlaGluPro----- 407
 DB 1731 GATTTATCCAAAGTTGGCTGAACCTGAATGGGAGATACCATCTTTGATAGAAAGTAGATA 1790
 QY 407 ----- 407
 DB 1791 GATTATTTCAAAGTAGGAGTAGCTTTGTCTTACAAAAGCATTTGCCTCTGATACAGAAAAG 1850
 QY 407 ----- 407
 DB 1851 AAGATGTTTGAAGCATTTCAATAGTATTAATCTGCGGGAACGAAATGAACCATCAGAG 1910
 QY 408 -----TriLeuThrPheGly 412
 DB 1911 TTACTTGTGATGAAATACTTACAACTTTAGCAAGCAGACGAATTTGTAATTTTGCA 1970
 QY 413 ValGlnProAsnProProValProSerProLysLeuGlyTrpMetAspGluAla----- 430
 DB 1971 ACAAGTAACAACATTTGCCATCAGAGGTACATACCTGTTGTCATGAGCAACACCCGAC 2030
 QY 430 ----- 430
 DB 2031 TGGTTTTTCAAGGACACAAATGGAATACGTTGAGCAAGGATGCATTCCTAAGCAGATTA 2090
 QY 430 ----- 430

Db 2091 AAACAGTATATTATACGCTAGTGGGAAGATATAAAGGGAAGGTTTATGATGGGATGTG 2150
Qy 431 -----MetAlaIleAspProPheAsnSerAspArgMetLeu----- 442
Db 2151 GTAAATGAAGCAATAGATGAAGTCAAGGTGATGGATTGAGGAGATCTAACTGGTACAAC 2210
Qy 443 -----TyrGlyThrGlyAlaThrLeuTyrAla----- 451
Db 2211 ATTTGTAGTCCCGAATATATTGAGAAGGCTTTTATATGGGCACATGAAGCCGATCCAGAC 2270
Qy 452 -----ThrAsnAspLeuThrLysTyrAspSerGlyGlnIleHisIleAla 467
Db 2271 GCAAAATGTTTACCAACGATTACAACACAGAAAACAGTCAGAAAGACAGTTTATTATTC 2330
Qy 468 ProMetValIleSerLeuGluThrAlaValAsnAspLeuIleSerProPheSerGly 487
Db 2331 AACATGATTAAAGTCTCAAGGAAAAGGTGTT-----CCAATTCTATGGA 2375
Qy 488 AlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAla 507
Db 2376 ATAGGATTGCAGAGTCATATA-----AATCTTGATTGGCCCTCGATTAGCGAG 2423
Qy 508 ValProSerThrIle-----PheThrSerProValPheThrThrGlyThrSerValAsp 525
Db 2424 ATAGAACACCACTAAGATTGTTTCAGCTCT-----ATACCTGGATTGGAGATACAC 2474
Qy 526 TyrAlaGluLeuAsnProSerIleIleValAlaGlySerPheAspProSerSerGln 545
Db 2475 ATTACGAGCTTGATGATGATGTTTATCATGTCAGTGGGTTCGAGTACCAGTTACTCAACGCCA 2534
Qy 546 ProAsnAsp----- 548
Db 2535 CCAAGAGATCTCTGTATAAAACAGGCAATGAGATATAAAGGAGTTATTGATTGTTTAA 2594
Qy 549 -----ArgHisValAlaPhe----- 553
Db 2595 AAGTACAACAATGTATAACAAGTGTAACATTTCTGGGACTGAAGGATGATTACTCATGG 2654
Qy 554 ---SerThrAspGlyGlyLysAsn----- 560
Db 2655 CTGAGTCAAAACCTTGGAAAAGTGATGATTACCCGTTGTTATTGATGAAACTATAATCA 2714
Qy 561 -----TyrPheGlnGlySerGluProGlyValThrThrGlyGlyThrValAla 577
Db 2715 AAATATGCTTTTGGAGCCTGATTGAGCAACTGTGATACCGGCCAACTCAACATTCGCCA 2774
Qy 578 Ala-----SerAlaAspGlySerArgPheVal 586
Db 2775 GCACCACCACTATTCAAATACCTACACCAACTCCACACACCAACCCGACACCGACAGTG 2834
Qy 587 TrpAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTyr 606
Db 2835 AGTGCAACGCCAACACAGCA-----CCGACGGCATCCCGTAGGT---GGCAGTTACTGG 2888
Qy 607 AlaAlaSerGlnGlyValProAla-----AsnAlaGlnIle 618
Db 2889 ACGCCGAGTGAGAGTTACAGTGGCTGAGAGTATGTTGCGAATGGGAATTTAAGCAGC 2948
Qy 619 ArgSerAspArgValAsnProLysThrPheTyrAlaLeuSerAsn-----GlyThr----- 635
Db 2949 CCGACCAATGATTGAATCTTAAG-----ATAAAGATAGAGAATGTTGGGACGACACGCG 3002
Qy 636 -----PheTyrArgSerThrAspGlyGlyValThr----- 645
Db 3003 GTAGATCTTTAGCAGGTGAAGGTAAAGATACTGTTACACGATAGATGTTGGGCAACACAG 3062
Qy 646 -----PheGlnPro-----ValAlaAla 651
Db 3063 AGTGTAAGTGTAAACAGCAGCATAGATCTCTGCGTATATAGATGTGAAGTTTGAAGCTT 3122
Qy 652 GlyLeuProSerSerGlyAla-----ValGly 660
Db 3123 GGAGCGAACCGCAGCGGAGCGGATTACTATGTGAGATAGGCTTTAAGAGTGGAGCAGCG 3182

Qy 661 ValMetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGly 680
Db 3183 GTTTTGGCAGCAGCGCAAGCAGCAGAGATAGACTTAGCATACAGAAAGCGCAGTGGC 3242
Qy 681 LeuTyrHisSerThrAsnGlyGlySerSerTrpSerAla----- 693
Db 3243 AGCTACAAATCAGTCAAAATGACTATTCCGTGAGGAGTGCACAGGCTATATAGAACGAG 3302
Qy 694 ---IleThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGlySer 712
Db 3303 AAGTACACGGGTATATAGATGATGTTCTTATCGGGAAGAGACCGGACGAGAACGCC 3362
Qy 713 SerTyrProAlaValPheValGlyThrIleGlyGlyValThrGlyAla 729
Db 3363 CAGATCAAGGTATGTTGATCGAATGGGAATTTAAGCAGCCCGCAATGTTATTGAATCCT 3422
Qy 730 ---TyrArgSerAspAspCysGlyThrThr----- 738
Db 3423 AAGATAAAGATAGAGAAATGTTGGACGACAGCGGTAGATCTTTAGCAGGGTGAAGTAAAGA 3482
Qy 739 ---TrpValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThr 757
Db 3483 TACTGGTACACCATAGATGTTGGAGCAACACAGAGTGAAGTGAACAGCAGCATAAAC 3542
Qy 758 GlyAspHisAlaAsnLeuArgArgValTyrIleGlyThrAsnGlyArgGlyIle---Val 776
Db 3543 CCTCGTATATAGATGTGAAGTGTGAGAGCTTGGAGCAAAATGCAAGTGGAGCGGATTAC 3602
Qy 777 TyrGlyAspIleGly-----GlyAla----- 783
Db 3603 TATGTGAGATAGAGCTTTAAGAGTGGAGCGGGTTTGGCAGCAGGGCAGACGACGAAG 3662
Qy 784 -----ProSerGlySer----- 787
Db 3663 GAGATAAGACTTAGCATACAGAGGCGAGTGGCAGCTACAAATCAGTCAATGACTATTTCG 3722
Qy 787 ----- 787
Db 3723 GTGAGGAGTCAACAGGCTATATAGAGAACGAGAAGGTAAACGGGTATATAGATGTGCG 3782
Qy 788 -----ProSerProSerValSerProSerAlaSerProSerLeuSer 801
Db 3783 ATAGTGTGGGAAGAGAGCCGAGCGGGGTACAAAGCCGCGGAGTAGTAGAACCCGACA 3842
Qy 802 ProSerProSerProSerSerProSerProSerProSerProSerProSerProSer 821
Db 3843 CCGCACCGACCCCGACATCGACCGCGACCAACACCTTACCAACCACTGCACCGACA 3902
Qy 822 SerSerProSerProSerProSerProSerProSerProSerProSerProSerProSer 841
Db 3903 TCAGCCCGACACCGAGCCCAACAGTGAACAGCCGACTCCCAACCGCGACCGACA 3962
Qy 842 AlaSerProSerProSerProSerProSerProSerProSerProSerProSerProSer 861
Db 3963 GTGACGGTTACTGTGACTCCGACACCGACACCA-----ACACCGACG 4004
Qy 862 ProThrProSerSerProValSerGly---GlyValIleValGlnTyrIleValAsn 880
Db 4005 CCGCACCGACAGGACACCTGGCAGCGGAAAGTGGTTTGAAGGCTACTATACAAAGAACAT 4064
Qy 881 AspSerAlaProGlyAspAsnGlnIleLysProGlyLeuGlnValValAsnThrGlySer 900
Db 4065 GAGACAAGTGGCAGCACCAAGTTCTATAAGCGCGTGGTTAAGATAGTAGTAGGAGCAGC 4124
Qy 901 SerSerValAspLeuSerThrValThrValArgTyrTrpPheThrArgAspGlyGlySer 920
Db 4125 AGCAGTGTGTATCTTAGCAGGTTAAGATAAGATACTGGTACACAGTGGATGGTGACAAG 4184
Qy 921 SerThrLeuValTyrAsnCysAspTrpAlaIleGlyCysGlyAsnIleArgAlaSer 940
Db 4185 CCACAGAGTGGCGTA---TGTGACTGGGCACAGATAGGGGCAAGCAATGTGACTTCAAT 4241

QY 941 PheGlySerValIleProAlaThrProThrAlaAspThrTyrLeuGln 956
 Db 4242 TTTGTGAAGCTGAGCGGAGTGAGTGGAGCGGATTATTACTTTGGAG 4289

RESULT 11

AAD26525
 ID AAD26525 standard; DNA; 11707 BP.

XX AC AAD26525;

XX DT 26-MAR-2002 (first entry)

XX DE Active cellulase protein, celB gene.

XX KW Active cellulase protein; alkalophilic; textile processing; proteinase;
 KW detergent additive; stonewashed appearance; cotton-containing denim;
 KW CelB5; thermophilic; commercial detergent; celB gene; ds.
 XX Unidentified.

XX OS

XX FH Key

XX CDS 6255..10535

XX FT Location/Qualifiers

XX FT /*tag= a

XX FT /product= "CelB protein"

XX FT 8601..10532

XX FT /*tag= b

XX FT /product= "DNA encoding B4/5 protein"

XX FT 9255..10526

XX FT /*tag= c

XX FT /product= "DNA encoding B/5 protein"

XX FT

XX PN US6294366-B1.

XX PD 25-SEP-2001.

XX PF 19-AUG-1998; 98US-00136574.

XX PR 19-SEP-1997; 97US-00932571.

XX PA (CLRN) CLARIANT FINANCE BVI LTD.

XX PI Farrington GK, Anderson P, Bergquist P, Daniels R, Gibbs MD;

XX PI Morgan H, Williams DP;

XX XX WPI; 2002-081780/11.

XX DR P-PSDB; AAE16323.

XX New cellulase active protein, useful in textile processing or commercial
 PT detergents, e.g. for improving the feel or appearance of cotton-
 PT containing fabrics, is stable under conditions of alkaline pH and
 PT elevated temperatures.

XX PS Disclosure; Col 27-38; 61pp; English.

XX The present invention relates to a cellulase active protein, which is
 CC substantially free of proteinases of native thermophilic and
 CC alkalophilic origin, where the cellulase active protein consists of the
 CC CelB5 amino acid sequence. The cellulase active protein is useful for
 CC treating cellulosic materials including cotton-containing fabrics, as
 CC detergent additives. The cellulase active protein is also useful for
 CC improving the feel and/or appearance of cotton-containing fabrics, for
 CC removing surface fibers from cotton-containing knits or for imparting
 CC stonewashed appearance to cotton-containing denims. The present proteins
 CC are stable under condition of alkaline pH and elevated temperatures, thus
 CC suitable for textile processing and in commercial detergents. The present
 CC sequence is celB gene

XX SQ Sequence 11707 BP; 3827 A; 1995 C; 2994 G; 2891 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 9.59e-06 Length: 11707
 Score: 409.00 Matches: 263
 Percent Similarity: 32.37% Conservative: 163

Best Local Similarity: 19.98% Mismatches: 386
 Query Match: 7.96% Indels: 504
 Db: 6 Gaps: 58

US-09-917-376-1 (1-957) x AAD26525 (1-11707)

QY 14 ArgArgLeuValSerLeuLeuAlaAlaThrAlaSerPheAlaValAlaAlaLeuGly 33

Db 735 AAGAGGGTAATTCAAATCTT---TCCTTATTGTTTTTTTAAATAAACACGCTGTGAGT 791

QY 34 ValLeuProIleAlaIleThrAlaSerProAlaHisAlaAlaThrGlnProTyrThr 53

Db 792 ACTTGT-----ATATTCATCAGGAAGCAAAGCAGCAGCA-----TATACT 833

QY 54 TrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsnGluGly 73

Db 834 -----GTTGATTTGAGGT 848

QY 74 AlaProGlyIleLeuTyr-----ValArgThrAspIleGlyGly 86

Db 849 GCTGATACATTTATCTTACTTTGCTTATGGAATAATCGAGCATAGCAGTTGACATGGCAAT 908

QY 87 MetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrp 106

Db 908 ----- 908

QY 107 AsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLys 126

Db 909 -----GCATATAATGGTAAAGTAGTGTGAGGGTGTCA-----AATAGAAGTTCA 953

QY 127 ValTrpAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu 146

Db 954 ATATGGGATGAGTTGCAGTT-----CACGTTAA 983

QY 147 ArgSerSerAspGlnGlyAlaThrTrpGlnIleThr----- 158

Db 984 AACATTATGAACAATGGAAACCACTAGTGTTCAGCGTATGTAATAACATAGTACCAG 1043

QY 159 ---ProLeuProPheLysLeu-----GlyGlyAsnMetProGlyArg 171

Db 1044 AAGCCGTTGCAATTTGGTATCTCAGCGGTTTACGACGATGGAGTGGGTAAAGAGTACT 1103

QY 172 GlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAla 191

Db 1104 CTCATAGTGAAGTGTGCGCTATT---CCAAAT----- 1133

QY 192 ProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThr 211

Db 1134 -----TATTGGAAGAAAAATT--- 1148

QY 212 AsnPheProAspValGlyTyrTyrIleAlaAsnProThrAspThrGlyTyrGlnSer 231

Db 1149 -----GTTGGTAAATGGACTCCAATATATGCAATGTGAGCAATTTGTTAATT 1196

QY 232 AspileGlnGlyValValTrpValAlaPheAspLysSerSerSerSerLeuGlyGlnAla 251

Db 1197 GTAATACACACAAATTTGTAAGCAAGCAAGTATGATAATGTTGAC----- 1241

QY 252 SerLysThrIlePheValGlyValAlaAspProAsn-----AsnProValPhe 267

Db 1242 -----TATATCCAATAATGATGATAATAGTTACCTATCAATGCAAGTACGA 1289

QY 268 TrpSerArg-----AspGlyGlyAlaThr-----TrpGlnAla-----Val 279

Db 1290 TTTTCAAGTGGATTGAAAGTGCGCACTACCGAGGGTTGCGAGGCAAGGGAAGCGGTGT 1349

QY 280 ProGlyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsnHisVal 299

Db 1350 ACAGTAAACACAGATAGCGTTGTGCGCATATATAGTGGCAAGTATAGT----- 1394

QY 300 LeuTyrIleAla-----ThrSerAsnThrGlyGly-----ProTyrAspGly--- 313

Db 1395 TTGTACGTGAGTGAAGAACCGTCAATTTGGCATGTTGGTGCACAGATTCGCGTAGATCAATT 1454

Qy	314	---SerSerGlyAspValTrpLysPheSerVal-----ThrSerGlyThrTrp	328
Db	1455	TTGGAAACAGGCTAAAGTGTATAAAATAAGTGTGGTATTATCAGAACAGCTGCTCAACT	1514
Qy	329	ThrArgIleSerProValProSerThrAspThrAlaAsnAsp---TyrPheGlyTyrSer	347
Db	1515	CAAAAATATGCTTAACACTGCAAGAAGAATTTGCTACAGATCCTTCAACAAGACTAGAA	1574
Qy	348	GlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrp	367
Db	1575	AACTCTGATATATACAGCGATGTACCGAGTAATACGTGGGTTCAGCTGAGTGAAAGCTAC	1634
Qy	368	TrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIleTrp	387
Db	1635	-----TCAAATCTCTGCTGCTTACAGTTAGCGAGTGGTGTG	1670
Qy	388	AspTrpThrSerTyrProAsnArgSerLeuArgTyr-----ValLeu	401
Db	1671	CTTTATGTTGAGGCACAAAATGCAAAATTTGGCTTCTTGGGTTCATGATTTAAAGATTTAT	1730
Qy	402	AspIleSer-----AlaGluPro-----	407
Db	1731	GAITATCCAAAGTTGGCTGAACTGAACTGAATGGGAGATACCATCTTTGATAGAAAAGTATAGA	1790
Qy	407	-----	407
Db	1791	GAITATTTCAAAGTAGGATAGCTTTGTCTTACAAAAGCATTGCCTCTGATACAGAAAAG	1850
Qy	407	-----	407
Db	1851	AAGATGTTTTTGAAGCATTTTCAATAGTATTACTGTCAGGGAACGAAATGAAACCATCAGAG	1910
Qy	408	-----	408
Db	1911	TTACTTGTGCGATGAAATACATTCAACTTTAGCAAAGCAGACGCAATTTGAAATTTTGCA	1970
Qy	413	ValGlnProAsnProProValProSerProLysLeuGlyTrpMetAspGluAla-----	430
Db	1971	ACAAGTAACAACATGGCCATCAGAGGTCATACACTGGTTGGCATGAGCAACACCCGAC	2030
Qy	430	-----	430
Db	2031	TGTTTTTTTCAAGGACACAAATGGAATACGTTGAGCAAGGATGCAITGCTAAGCAGATTA	2090
Qy	430	-----	430
Db	2091	AAACAGTATATTTATACGGTAGTGGGAAGATATAAAGGGAAGTTTATGATCGGATGTG	2150
Qy	431	-----MetAlaIleAspProPheAsnSerAspArgMetLeu-----	442
Db	2151	GTAATGAAGCATAGATGAAGTCAAGGTGATGGATTTCAGGAGATCTAACTGGTACAC	2210
Qy	443	-----TyrGlyThrGlyAlaThrLeuTyrAla-----	451
Db	2211	ATTGTAGTCCCGAATATATTGAGAAGGCTTTTATATGGGCACATCAAGCCGATCCAGAC	2270
Qy	452	-----ThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAla	467
Db	2271	GCAAAATGTTTATCAACAGTATCAACACAGAAACACAGTCAGAGACAGACAGTTATTATC	2330
Qy	468	ProMetValLysGlyLeuGluGluThrAlaValAsnAspLeuIleSerProProSerGly	487
Db	2331	AACATGATTAAAGTCTCAAGGAAAAAGGTGTT-----CCAATTCATGGA	2375
Qy	488	AlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAla	507
Db	2376	ATAGGATTCAGAGTCAATA-----AATCTTGATTGGCCCTCGATTAGCGAG	2423
Qy	508	ValProSerThrIle-----PheThrSerProValPheThrGlyThrGlyThrSerValAsp	525
Db	2424	ATAGAGAACACCATTAAGATTTGTTTCAGCTCT-----ATACCTGATTGGAGATACAC	2474

Qy	526	TyrAlaGluLeuAsnProSerIlelleValArgAlaGlySerPheAspProSerSerGln	545
Db	2475	ATTACGGAGCTTGTATATGAGTTTATTCAGTGGGGTTCGAGTACCAGTTACTCAACGCCA	2534
Qy	546	ProAsnAsp	548
Db	2535	CCAAGAGATCTCTGTATAAACAACGCGCATGAGATATAGGAGTTATTTGATTTGTTTAAA	2594
Qy	549	-----ArgHisValAlaPhe	553
Db	2595	AAGTACAACAATGTAATAAACAAGTGTAAACATCTCTGGGACTGAAGGATGATTACTCATGG	2654
Qy	554	---SerThrAspGlyGlyIysAsn	560
Db	2655	CTGAGTCAAAACTTTGGAAAAAGTATTACCCGTTGTTATTGATGAAAACTATAAATCA	2714
Qy	561	-----TrpPheGlnGlySerGluProGlyValThrGlyGlyThrValAla	577
Db	2715	AAATATGCTCTTTGGAGCTGATTGAGCCAACTGTGTATACGGCGCAACTCAACANTGCCA	2774
Qy	578	Ala	586
Db	2775	GCACCACGCTATTCAAATACCTACACCAACTCCACACCAACCCCGACACCGACAGTG	2834
Qy	587	TrpAlaProGlyAspProGlyGlnProValValTyAlaValGlyPheGlyAsnSerTrp	606
Db	2835	AGTGCAACGCCAACACACAGCA---CGACGGCATCACCGTAGGT---GGCAGTTACTGG	2888
Qy	607	AlaAlaSerGlnGlyValProAla	618
Db	2889	ACGCCGAGTGAGAGTTACAGTGGCGCTGAAGGTATGTTATGCGAATGGGAATTTAAGCAGC	2948
Qy	619	ArgSerAspArgValAsnProLysThrPheTyAlaLeuSerAsn	635
Db	2949	CCGACGAATGTATTGATTCCTAG-----ATAAGATAGAGAAATGTTGGACGACAGCG	3002
Qy	636	-----PheTyArgSerThrAspGlyGlyValThr	645
Db	3003	GTAGATCTTAGCAGGGTGAAGGTAAAGTACTGCTACACGATAGATGCTGAGGCAACACAG	3062
Qy	646	-----PheGlnPro	651
Db	3063	AGTGTAAGTGTAAACAGCACATAGATCTCTCGCTATATAGATGTGAAGTTGTGAAGCTT	3122
Qy	652	GlyLeuProSerSerGlyAla	660
Db	3123	GGACGAAACGACGGCGAGCGGATTACTATGTGGAGATAGGCTTTAAGAGTGGACAGGG	3182
Qy	661	ValMetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaSerSerGly	680
Db	3183	GTTTGGCAGCAGCGGCAACAGCACGAGGAGATAAGACTTAGCATACAGAGGGCAGTGGC	3242
Qy	681	LeuTyRHisSerThrAsnGlyIysSerSerTrpSerAla	693
Db	3243	AGCTACAATCAGTCAAAATGACTATTCGGTGAGAGTGCACACAGCTATATAGAGAACGAG	3302
Qy	694	--IleThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGlySer	712
Db	3303	AAGTAAACGGGTATATAGATGATGATGTTACTTGTATGGGAAGACAGCGCAGGAAACGCC	3362
Qy	713	SerTyRProAlaValPheValValGlyThrIleGlyValThrGlyAla	729
Db	3363	CAGATCAAGGTATGTTATGCGAATGGGAATTTAAGCAGCCCGACGAATGTTATGAACTCT	3422
Qy	730	---TyrArgSerAspAspCysGlyThrThr	738
Db	3423	AAGATAAAGATAGAGAATGTTGGGACGACACAGCGGTAGATCTTAGCAGGGTGAAGGTAAAG	3482
Qy	739	---TrpValLeuIleAsnAspAspGlnHisGlyTyrGlyAsnTrpGlyGlnAlaIleThr	757
Db	3483	TACTGGGTACAGATAGATGTGTGAGGCAACACAGAGTGAAGTGTAAACAGCAGCATTAAC	3542
Qy	758	GlyAspHisAlaAsnLeuAqArgValTyIleGlyThrAsnGlyArgGlyIle---Val	776

```
Db 3543 CTTGGCTATATAGATGTAAGATTGTAAGCTTGGACCAATGCGAGCGGATTAC 3602
Qy TyrGlyAspIleGly-----GlyVala----- 783
Db 3603 TATGTGGAGATAGCTTTAAGATGGAGCAGGGGTTTGGCAGCAGGGCAGACGCAAG 3662
Qy 784 -----ProSerGlySer----- 787
Db 3663 GAGATAAGACTTAGCATACAGAAGGCGAGTGGCAGCTACAATCAGTCAATGACTATTTCG 3722
Qy 787 ----- 787
Db 3723 GTGAGGAGTCAACAGCTATATAGAACGAGAAGGTAAACGGGTATATAGATGGCG 3782
Qy 788 -----ProSerProSerValSerProSerAlaSerProSerLeuSer 801
Db 3783 ATAGTGGGAGAGAGAGCCGAGCGGCTACAAAGCCGGGGAGTAGTAACACCGACA 3842
Qy 802 ProSerProSerProSerProSerProSerProSerProSerProSerProSer 821
Db 3843 CCGGACCGACCCGAGATCGACCGGACACCAACCTTACACCACTGCACCGACA 3902
Qy 822 SerSerProSerProSerProSerProSerProSerProSerProSerProSer 841
Db 3903 TCAGCCCGACACCGAGCCCAACAGTACAGCAACGCCGACTCCACCGCGAGCGCGACA 3962
Qy 842 AlaSerProSerProSerProSerProSerProSerProSerProSerProSer 861
Db 3963 GTACGGTGTACTGTGACTCCGACACCGACCA-----ACACCGAGC 4004
Qy 862 ProThrProSerSerProValSerGly---GlyValLysValGlnTyrLysAsnAsn 880
Db 4005 CCACACCGACAGGACACCTGCGACCGGAGTGGTTGAAGGTACTATACAGAACAT 4064
Qy 881 AspSerAlaProGlyAspAsnGlnLysProGlyLeuGlnValValAsnThrGlySer 900
Db 4065 GAGACAAAGTCGAGCACAAGTCTTATAAGCCCGTGGTTAAGTAGTAGTGAAGCGCAGC 4124
Qy 901 SerSerValAspLeuSerThrValThrValArgTyrTrpPheThrArgAspGlyGlySer 920
Db 4125 AGCAGTGTGATCTTAGCAGGGTTAAGATAAGATATCTGTGTACACAGTGGTGTGACAAG 4184
Qy 921 SerThrLeuValTyrAsnCysAspTTrpAlaAlaIleGlyCysGlyAsnIleArgAlaSer 940
Db 4185 CCACAGAGTCGGTA---TGTGACTGGGCACAGATAGGGGCAAGCAATGTGACATCAAT 4241
Qy 941 PheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 956
Db 4242 TTTGTGAGCTGAGCAGCGAGTGGAGTGGAGCGGATATTACTTTGGAG 4289
RESULT 12
AAF14988
ID AAF14988 standard; cDNA; 1103 BP.
XX
AC AAF14988;
XX
XX 15-SEP-2003 (revised)
DT 13-MAR-2001 (first entry)
XX
XX Trichoderma reesei EST SEQ ID NO:7511.
XX
KW Multiple gene expression; filamentous fungal cell; EST;
KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW culture condition; environmental stress; spore morphogenesis;
KW metabolic pathway engineering; catabolic pathway engineering; ss.
XX
OS Hypocrea jecorina.
XX
XX WO200056762-A2.
XX
XX 28-SEP-2000.
```

```
XX 22-MAR-2000; 2000WO-US007781.
XX
XX 22-MAR-1999; 99US-00273623.
XX (NOVO ) NOVO NORDISK BIOTECH INC.
XX (NOVO ) NOVO NORDISK AS.
XX
XX Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
XX WPI; 2000-594572/56.
XX
XX Monitoring differential expression of genes in filamentous fungal cells
XX PT uses fluorescence-labeled nucleic acids isolated from the cells and a
XX PT substrate of expressed sequence tags.
XX
XX Claim 89; Page 3034; 3161pp; English.
XX
XX The present invention describes a method for monitoring differential
XX CC expression of genes in a first filamentous fungal (FF) cell relative to
XX CC expression of the same genes in one or more second filamentous fungal
XX CC cells. The method uses fluorescence-labeled nucleic acids isolated from
XX CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
XX CC are used in the methods for monitoring differential expression of genes
XX CC in a first filamentous fungal (FF) cell relative to expression of the
XX CC same genes in one or more second filamentous fungal cells. Monitoring the
XX CC global expression of genes from FF cells allows the production potential
XX CC of the microorganisms to be improved. New genes may be discovered,
XX CC possible functions of unknown open reading frames can be identified and
XX CC gene copy number variation and stability can be monitored. The expression
XX CC of genes can be used to study how FF cells adapt to changes in culture
XX CC conditions, environmental stress, spore morphogenesis, recombination,
XX CC metabolic or catabolic pathway engineering. Using ESTs provides several
XX CC advantages over genomic or random cDNA clones including elimination of
XX CC redundancy as one spot on an array equals one gene or open reading frame,
XX CC and organization of the microarrays based on function of the gene
XX CC products to facilitate analysis of the results. AAF07478 to AAF11247
XX CC represents ESTs from Fusarium venenatum; AAF11248 to AAF11853 represents
XX CC ESTs from Aspergillus niger; AAF11854 to AAF14878 represents ESTs from
XX CC Aspergillus oryzae; and AAF14879 to AAF15337 represents ESTs from
XX CC Trichoderma reesei, which are all specifically claimed in the present
XX CC invention. (Updated on 15-SEP-2003 to standardise OS field)
XX
SQ Sequence 1103 BP; 230 A; 322 C; 307 G; 223 T; 0 U; 21 Other;

Alignment Scores:
Pred. No.: 5.08e-06 Length: 1103
Score: 392.50 Matches: 114
Percent Similarity: 45.05% Conservative: 36
Best Local Similarity: 34.23% Mismatches: 125
Query Match: 7.64% Indels: 58
DB: 3 Gaps: 10

US-09-917-376-1 (1-957) x AAF14988 (1-1103)
Qy 348 GlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrp 367
Db 3 GGCCTTGGCTCGATTGTCGAAAGCCAGGACCCCTTGTGTGTCTTGTGAACCTCTGG 62
Qy 368 TrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIleTrp 387
Db 63 TGGCCAGATGCTCAGCTGTTTGGTCGACCGACTCTGGGACACATGGAGCCGATCTGG 122
Qy 388 AspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluPro 407
Db 123 GCGTGGCGGAGCTATCCGACTGAGACCTTATTACTACAGCATCTCAACTCCCAAGCACCG 182
Qy 408 TrpLeuThrPheGly---ValGlnProAsnProProValProSer-----Pro 422
Db 183 TGGATCAAGAACAACACTTTATCGATGTGTGACGCGAGTCACCGTCGCGATCTNATCAAG 242
Qy 423 LysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeu 442
XX
```


QY 780 eGlyGlyAlaProSerGlySerProSerProSerValSerProSerAlaSerProSerLe 800
Db 304 -----CCATCACCACCATCACCACCATCTCCATCACCACCATCTCCATCTCC 255
QY 800 uSerProSerProSerProSerProSerProSerProSerProSerProSerPr 820
Db 254 ATCACCATCACCATCTCCATCACCATCTCCATCTCCATCTCCATCACCACCATCACC 195
QY 820 oSerSerProSerProSerProSerProSerProSerProSerProSerArgSerProSerPr 840
Db 194 ATCACCATCACCATCACCATCTCCATCACCATCTCCATCTCCATCACCACCATCACC 135
QY 840 oSerAlaSerProSerProSerProSerProSerProSerProSerProSerSerPr 860
Db 134 ATCTCCATCACCATCACCATCTCCATCACCATCTCCATCACCATCACCATCTCCATCACC 75
QY 860 oSerProThrProSerSerPro 868
Db 74 ATCACCATCTCCATCACCACCATCACCACCA 50

RESULT 14
AAF31109/c

ID AAF31109 standard; cDNA; 18596 BP.

XX AC AAF31109;

XX XX 27-APR-2001 (first entry)

XX DE Thymidylate synthase coding sequence.

XX XX Analyte-binding enzyme; analyte analysis; ss.

XX KW Homo sapiens.

XX OS WO200102600-A2.

XX PN 11-JAN-2001.

XX PD 30-JUN-2000; 2000WO-US018057.

XX PF 06-JUL-1999; 99US-00347878.

XX PR 06-DEC-1999; 99US-00457205.

XX XX (GEAT) GEN ATOMICS.

XX PA Yuan C;

XX PI WPI; 2001-071583/08.

XX DR Assaying method, useful for prognosis and diagnosis of disease, comprises
XX PT contacting sample with a mutant analyte-binding enzyme and detecting
XX PT binding.

XX PS Disclosure; Page; 187pp; English.

XX CC The present invention relates to a method for assaying an analyte in a
XX CC sample comprising: contacting the sample with a mutant analyte-binding
XX CC enzyme which has binding affinity for the analyte or an immediate analyte
XX CC enzymatic conversion product but has attenuated catalytic activity; and
XX CC detecting resulting binding. The method is useful in monitoring
XX CC biological systems/processes, or prognosis/diagnosis of disease caused by
XX CC imbalances of the analytes. The present sequence is a coding sequence
XX CC used in the present invention. Note: the present sequence is not shown in
XX CC the specification, but was from Genbank, using information given in the
XX CC specification

SQ Sequence 18596 BP; 4521 A; 3991 C; 4479 G; 5605 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 0.00253 Length: 18596
Score: 345.00 Matches: 71
Percent Similarity: 83.91% Conservative: 2
Best Local Similarity: 81.61% Mismatches: 13

Query Match:

DB: 6.72% Indels: 1

US-09-917-376-1 (1-957) x AAF31109 (1-18596)

QY 783 AlaProSerGlySerProSerProSerValSerProSerAlaSerProSerLeuSerPro 802

Db 10388 TCACCATCACCATCTCCATCTCCATCACCACCATCACCACCATCACCACCA 10329

QY 803 SerProSerProSerProSerProSerProSerProSerProSerProSerSer 822

Db 10328 TCACCATCTCCATCACCATCACCATCTCCATCACCACCATCACCATCTCCATCACC 10269

QY 823 SerProSerProSerProSerProSerProSerProSerProSerProSerAla 842

Db 10268 TCACCATCACCATCACCATCTCCATCACCACCATCACCACCATCACCACCA 10209

QY 843 SerProSerProSerProSerProSerProSerProSerProSerProSerSer 862

Db 10208 TCACCATCACCATCACCACCATCACCACCATCACCACCATCACCACCATCACC 10149

QY 862 oThrProSerSerPro 868

Db 10148 ATCTCCATCACCACCATCACCACCA 10130

RESULT 15
AAC91215/c

ID AAC91215 standard; DNA; 18596 BP.

XX AC AAC91215;

XX XX 20-MAR-2001 (first entry)

XX DE Human thymidylate synthase gene SEQ ID NO: 11.

XX KW Human; schizophrenia; developmental disorder; spina bifida cystica;
XX KW Tourette's syndrome; bipolar illness; autism; conduct disorder;
XX KW attention deficit disorder; obsessive compulsive disorder;
XX KW chronic multiple tic syndrome; learning disorder; polymorphism; ds.
XX OS Homo sapiens.
XX XX WO200071754-A1.
XX PD 30-NOV-2000.
XX PF 24-MAY-2000; 2000WO-US014354.
XX PR 25-MAY-1999; 99US-00318448.
XX XX (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.
XX PI Johnson WG, Stenroos ES;
XX DR WPI; 2001-025174/03.

XX PT Diagnosing a developmental disorder, e.g. schizophrenia, by forming
XX PT datasets (DS) of genetic (e.g. genotypes of folate metabolism alleles)
XX PT and environmental variables affecting an individual and then comparing
XX PT these DS with reference DS.
XX PS Disclosure; Page 125-131; 156pp; English.
XX CC The present invention provides a novel method of estimating the
XX CC susceptibility of an individual to a developmental disorder using genetic
XX CC and environmental variables. The method can be used in the diagnosis,
XX CC prevention and treatment of disorders such as schizophrenia, spina bifida
XX CC cystica, Tourette's syndrome, bipolar illness, autism, conduct disorders,
XX CC attention deficit disorder, obsessive compulsive disorder, chronic
XX CC multiple tic syndrome and learning disorders such as dyslexia

SQ Sequence 18596 BP; 4521 A; 3991 C; 4479 G; 5605 T; 0 U; 0 Other;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 14, 2004, 08:53:05 ; Search time 5623 Seconds
(without alignments)
5082.359 Million cell updates/sec

Title: US-09-917-376-1
Perfect score: 5135
Sequence: 1 MDRSENIRLTMRSRLVSL.....RASFGSVNPATPTADTYLQX 957

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DRV=xlh
-Q=/cgn2.1/USPTO.spool/US09917376/runat_14052004_085257_4630/app.query.fasta_1.1095
-DB=EST -QPMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEARSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09917376 @CGN_1_1_4087@runat_14052004_085257_4630 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pin:*
20: em_gss_vit:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	631.5	12.3	704	14	CF872104	trico29xo
2	631.5	12.3	761	14	CF902332	trico29xo
3	603.5	11.8	747	14	CF876916	trico074xd
4	603.5	11.8	814	14	CB905388	trico074xd
5	415.5	8.1	707	14	CF880713	trico082xn
6	415.5	8.1	782	14	CB907625	trico082xn
7	401.5	7.8	693	14	CF882065	trico029xo
8	362	7.0	748	29	CE342004	tigr-gss-
9	361.5	7.0	751	14	CF868882	trico16xm
10	361.5	7.0	803	14	CB898982	trico16xm
11	360	7.0	791	29	CNS028QT	Tetraodon
12	355	6.9	522	29	CE190185	tigr-gss-
13	353	6.9	493	28	CC066065	CSU-K33r.
14	350	6.8	694	28	CC119324	NDL-70K21
15	347	6.8	259	28	AZ374273	LM0126G21
16	347	6.8	587	13	C79507	Mous
17	338	6.6	738	14	CF866388	trico006xe
18	338	6.6	794	14	CB896354	trico006xe
19	333	6.5	703	29	CE176982	tigr-gss-
20	329.5	6.4	476	13	C79500	Mous
21	328	6.4	562	28	AZ855343	2M0159A22
22	314.5	6.1	546	10	BF072664	NCSM3H773
23	314	6.1	619	28	BZ192350	CH230-248
24	314	6.1	959	29	CNS022KT	Tetraodon
25	310	6.0	744	28	BZ277113	CH230-450
26	307	6.0	538	29	CC847893	NDL-4416
27	306.5	6.0	389	13	BUG39045	mgcwo11xd
28	306	6.0	961	29	CNS03QKT	AL256070 Tetraodon
29	301.5	5.9	464	28	AZ928926	479.dif19
30	298.5	5.8	440	28	AZ312209	LM0027L19
31	297	5.8	929	14	CF885920	trico085xf
32	292	5.7	393	28	AZ332252	LM0060H12
33	291	5.7	761	29	CNS041RC	Tetraodon
34	287	5.6	559	28	BH306809	CH230-100
35	286.5	5.6	713	14	CF875815	trico039xx
36	286.5	5.6	782	14	CB904767	trico039xx
37	285	5.6	534	28	AZ014305	RPC1-23-3
38	284	5.5	592	28	AZ015337	RPC1-23-3
39	284	5.5	632	29	CE696112	tigr-gss-
40	282.5	5.5	719	14	CF881775	trico085xf
41	282.5	5.5	775	14	CB908435	trico085xf
42	282	5.5	647	9	AU061620	AU061620
43	274	5.3	401	12	BG887959	EST513810
44	272.5	5.3	697	28	AZ374916	LM0128I06
45	270	5.3	400	28	AQ171567	HS_3163_B

ALIGNMENTS

RESULT 1	CF872104	LOCUS	trico29xo13.b1	T. reesei	mycelial culture, Version 6 October 2003	linear	EST 31-OCT-2003
DEFINITION	CF872104	Hypocrea jecorina	CDNA clone	trico29xo13,	mRNA sequence.		
ACCESSION	CF872104.1	GI:38126786					
VERSION	EST.						
KEYWORDS	Hypocrea jecorina (anamorph: Trichoderma reesei)						
SOURCE	Hypocrea jecorina						
ORGANISM	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;						
REFERENCE	1 (bases 1 to 704)						


```
Db 146 CATGCTGCC-----TTTTTCATGGAAGACGTCAAGCTCGCGCGCGCGCGG 193
Qy 64 PheValAspGlyLeuValPheAsnGluGlyAlaProGlyLeuTyrValArgThrAsp 83
Db 194 TTTCGTCCTCCCGCATCATCTTCCATCCCAAGCAAAAGGCGTAGCATATGCAACAGAT 253
Qy 84 IleGlyGlyMetTyrArg-----TTPAspAlaAlaAsnGlyArgTrpLeuPro-----Leu 100
Db 254 ATTGGCGGGGTACCGGCTCTTCTCCGCACTCATGGACCGCC-GTCCAGGATGGGATT 312
Qy 101 LeuAspTrpValGlyTyrPheAsnTrpGlyTyrAsnGlyValValSerIleAlaAsp 120
Db 313 GCTGATAAATGCCGCTGGCGCAACTGG-----GGCATCGACGCTGTGCGCTTGTAT 363
Qy 121 ProIleAsnThrAsnLysValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspPro 140
Db 364 CCGCAGGACGATCAAAAGGTGTATGCCGAGTCGGCATGTATACGAACAGCTGGGATCCG 423
Qy 141 AsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeu 160
Db 424 AGTAATGGAGCCATCATCTCGTCGTGACACCGCGCGCAACGTGGTCTTCCACCACTTG 483
Qy 161 ProPheIysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAsp 180
Db 484 CCCTTCAAAATCGGGGGTAAATGCCAGGACGCGGAGCGCGAGAGCGTCTGGCTGTGAT 543
Qy 181 ProAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTyrArgSer 200
Db 544 CCGGCCACTCCAAATCATCTACTTGTGTCTGCTCAGGAAACGGCTCTCGAAGTCT 603
Qy 201 ThrAspSerGlyAlaThrTrpSerGlnMetMetAsnPheProAspValGlyThrTrpIle 220
Db 604 ACGGACGGCGGCGTGACCTTTCCAAAGTCTCGTCTGTTACGGCAACTGGAGCTATCATC 663
Qy 221 AlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAla 240
Db 664 CCAGACCCGAGTGATTCACACGGCTACCAACAGGATCAGCAGGATCATGTGGGTACG 723
Qy 241 PheAspLysSerSerSerLeuGlyGlnAlaSer 252
Db 724 TTCGACTC-ACCAGCAGCAGCAGCGGNGGAGCCACG 758
```

```
RESULT 3
CF876916 747 bp mRNA linear EST 31-OCT-2003
LOCUS trico074xd02.b1 T.reesei mycelial culture, version 6 October 2003
DEFINITION Hypocrea jecorina cDNA clone trico074xd02, mRNA sequence.
ACCESSION CF876916
VERSION CF876916.1 GI:38131598
KEYWORDS EST.
SOURCE Hypocrea jecorina (anamorph: Trichoderma reesei)
ORGANISM Hypocrea jecorina
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
REFERENCE 1 (bases 1 to 747)
AUTHORS Diener S.E., Dankmeyer L., Dunn-Coleman N., Houfek T.D.,
Mitchell, T.K., van Solingen, P., Teunissen, P.J.M., Ward, M. and
Dean, R.A.
TITLE Analysis of the protein processing and secretion pathways in a
Trichoderma reesei EST dataset
JOURNAL Unpublished (2003)
COMMENT Contact: Ralph A. Dean
Fungal Genomics Laboratory
North Carolina State University
Campus Box 7251, Raleigh, NC 27695, USA
Tel: 919-513-0020
Fax: 919-513-0024
Email: ralph.dean@ncsu.edu
Seq primer: Tr-F1 primer.
Location/Qualifiers
1..747
/organism="Hypocrea jecorina"
/mol_type="mRNA"
```

```
/strain="QM6a"
/db_xref="taxon:51453"
/clones="trico074xd02"
/dev_stage="mycelia"
/clone_lib="T.reesei mycelial culture, Version 6 October 2003"
/note="vector: pREP3; Site_1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."
```

Alignment Scores: 4.43e-13 Length: 747
Pred. No.: 603.50 Matches: 121
Score: 65.43% Conservative: 38
Percent Similarity: 49.79% Mismatches: 64
Best Local Similarity: 11.75% Indels: 20
Query Match: 14 Gaps: 6
DB:

```
US-09-917-376-1 (1-957) x CF876916 (1-747)
Qy 25 SerPheAlaValAlaAlaLeuGlyValLeuProIleAlaIleThrAlaSerProAla 44
Db 61 TCTCGATCTCTTGCCTTGTCTGGGCGCCGTC-----ATCCTGCCC 102
Qy 45 HisAlaAlaThrThrGlnProTyrTrpSerAsnValAlaIle---GlyGlyGlyGly 63
Db 103 CATGCTGCC-----TTTTTCATGGAAGAACGTCAAGCTCGCGCGCGCGCGCGC 150
Qy 64 PheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAsp 83
Db 151 TTGTCCTCCCGCATCATCTTCCATCCCAAGCAAAAGCGGTAGCATATGACGAACAGAT 210
Qy 84 IleGlyGlyMetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrp 103
Db 211 ATTGGCGGGGTGTACCGCCTCAAC---GCCAGCACTCATGGACCGCGCTCAGGATGGG 267
Qy 104 Val-----GlyTrpAsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAla 119
Db 268 ATTGCTGATTAATGCGGCTGGCACAACTGG-----GGCATCGACGCTGTGGGCTT 318
Qy 120 AspProIleAsnThrAsnLysValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAsp 139
Db 319 GATCCGCGACGACGATCAAAAGGTGTATGCGCAGTCGCGCATGTATACGAACAGCTGGGAT 378
Qy 140 ProAsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThrPro 159
Db 379 CCGAGTAATGGAGCCATCATCTCGCTCGTCAGACCGCGCGCAACGTGGTCTTCCACCAAC 438
Qy 160 LeuProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaVal 179
Db 439 TTGCGCTTCACATCGGGGGTAAATGCCAGAGACCGGAGCGGAGAGCGGTCTGGCTGTC 498
Qy 180 AspProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArg 199
Db 499 GATCCGCGCACCTCCAAACATCATCTACTTGTGTGCTCGCTCAGCAGACGGCTCTGGAAG 558
Qy 200 SerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyr 219
Db 559 TCTACCGACGCGGCGGTGACCTTTTCCAAAGCGCTCGTCGCTCAGCGCCCTGGGACGTC 618
Qy 220 IleAlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTrpVal 239
Db 619 CTTCCAGACCCGAGTGTATCCACGGGTACCAACCGCGACAGCAAGCGGCTCATGTGGGT 678
Qy 240 AlaPheAspLysSerSerSerSerLeuGly-GlnAlaSerLysThrIlePheValGlyVa 259
Db 679 ACGTTCGACTCACCCAGCATCACACCGGTGGCAGCCAGCGCTCGTATTTTGTGTCGCCAC 738
Qy 259 lAlaAsp 261
Db 739 GACTGAC 745
```

```

RESULT 4
CB905388      814 bp  mRNA  linear  EST 02-JUL-2003
LOCUS         trico074xd02 T.reesei mycelial culture, Version 3 april Hypocrea
DEFINITION   jecorina cDNA clone trico074xd02, mRNA sequence.
ACCESSION    CB905388
VERSION      CB905388.1 GI:30120046
KEYWORDS     EST.
SOURCE       Hypocrea jecorina (anamorph: Trichoderma reesei)
ORGANISM     Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
              Hypocreales; Hypocreales; Hypocreaceae; Hypocrea.
REFERENCE    1 (bases 1 to 814)
AUTHORS      Foreman,P.K., Brown,D.E., Dankmeyer,L., Dean,R., Diener,S.,
              Dunn-Coleman,N.S., Goedegebuur,F., Houfek,T.D., England,G.J.,
              Kelley,A.S., Meeran,H.J., Mitchell,T., Mitchinson,C.,
              Olivares,H.A., Teunissen,P.J., Yao,J. and Ward,M.
TITLE        Transcriptional regulation of biomass-degrading enzymes in the
              filamentous fungus Trichoderma reesei
JOURNAL      J. Biol. Chem. 278 (34), 31988-31997 (2003)
MEDLINE      22803314
PUBMED       12788920
COMMENT      Contact: Pamela K. Foreman
              Genencor Intl.
              925 Page Mill Road, Palo Alto, CA 94304, USA
              Tel: (650) 846-7635
              Fax: (650) 621-7817
              Email: Pforeman@genencor.com
              Seq primer: LT-P1 primer.
              Location/Qualifiers
                1..814
                  /organism="Hypocrea jecorina"
                  /mol_type="mRNA"
                  /strain="QM6a"
                  /db_xref="taxon:51453"
                  /clone="trico074xd02"
                  /dev_stage="mycelial"
                  /notes="Vector: pREP3Y; Site 1: Not I/Sal I; Mycelial
                  culture grown from 24 hrs to 6 days with varying Carbon
                  and Nitrogen sources and concentrations."

ALIGNMENT SCORES:
Pred. No.: 4,85e-13 Length: 814
Score: 603.50 Matches: 121
Percent Similarity: 65.43% Conservative: 38
Best Local Similarity: 49.79% Mismatches: 64
Query Match: 11.75% Indels: 20
DB: 14 Gaps: 6

US-09-917-376-1 (1-957) x CB905388 (1-814)

QY 25 SerPheAlaValAlaAlaLeuGlyValLeuProIleAlaIleThrAlaSerProAla 44
Db 128 TCTCGAGTCCTTGCCTTGTCTGGGGCGGTC-----ATCCCTGCC 169
QY 45 HisAlaAlaThrThrGlnProTyrThrTTPSerAsnValAlaIle---GlyGlyGly 63
Db 170 CARGCTGCC-----TTTTCATGAAGAAGACGTCAAGTCGCGCGCGCGCGC 217
QY 64 PheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAsp 83
Db 218 TTGCTCCCGGATCATCTTCATCCCAACAAAGGCGTAGCATATGCAGACAGAT 277
QY 84 IleGlyGlyMetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrp 103
Db 278 ATTGGGGGCTGTACCGCTCAAC---GCCGACGACTCATGGACCGCGCGTCACGATGG 334
QY 104 Val-----GlyTrpAsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAla 119
Db 335 ATTGCTGATAATGCCGCGTGGCAACTGG-----GCCATCGAGCTGTTCGCTT 385

```

```

QY 120 AspProIleAsnThrAsnLysValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAsp 139
Db 386 GATCCGCGAGCATGATCAAAAGGTGTATCCGCGCATGTATACGAACAGCTGGGAT 445
QY 140 ProAsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThrPro 159
Db 446 CCGAGTAATGGAGCCATCATTCGTCGTACAGCCGGCGGCAACGTGGTCTTCACCAAC 505
QY 160 LeuProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaVal 179
Db 506 TTGCGCTTCACAGTCGGGGTTAATCATGCCAGGACCGGAGCGGAGCGTCTGGCTGTC 565
QY 180 AspProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArg 199
Db 566 GATCCGCGCCACCTCAACATCATCTACTTTGGTGTCTCGCTCACGACACGGGCTCTGGAAG 625
QY 200 SerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyr 219
Db 626 TCTACCGACGGCGCGGTGACCTTTTCAGGCGCTGCTCGCTCACGGCCCTGGGAGTCC 685
QY 220 IleAlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTrpVal 239
Db 686 CTCACGACCGCGAGTGATTCACCGCTACCAACCGGACCAAGCAAGGCTCATGTGGGTT 745
QY 240 AlaPheAspLysSerSerSerSerLeuGly-GlnAlaSerLysThrIlePheValGlyVa 259
Db 746 ACGTTGACTCACCCAGCATCACGACCGGTGGCCACCGCTCGTATTTTGTCTGCCAC 805
QY 259 LalaAsp 261
Db 806 GACTGAC 812

RESULT 5
CF880713      707 bp  mRNA  linear  EST 31-OCT-2003
LOCUS         trico82xn19.b1 T.reesei mycelial culture, Version 6 October 2003
DEFINITION   Hypocrea jecorina cDNA clone trico82xn19, mRNA sequence.
ACCESSION    CF880713
VERSION      CF880713.1 GI:38135395
KEYWORDS     EST.
SOURCE       Hypocrea jecorina (anamorph: Trichoderma reesei)
ORGANISM     Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
              Hypocreales; Hypocreales; Hypocreaceae; Hypocrea.
REFERENCE    1 (bases 1 to 707)
AUTHORS      Diener,S.E., Dankmeyer,L., Dunn-Coleman,N., Houfek,T.D.,
              Mitchell,T.K., van Solingen,P., Teunissen,P.J.M., Ward,M. and
              Dean,R.A.
TITLE        Analysis of the protein processing and secretion pathways in a
              Trichoderma reesei EST dataset
JOURNAL      Unpublished (2003)
COMMENT      Contact: Ralph A. Dean
              Fungal Genomics Laboratory
              North Carolina State University
              Campus Box 7251, Raleigh, NC 27695, USA
              Tel: 919-513-0020
              Fax: 919-513-0024
              Email: ralph.dean@ncsu.edu
              Seq primer: LT-F1 primer.
              Location/Qualifiers
                1..707
                  /organism="Hypocrea jecorina"
                  /mol_type="mRNA"
                  /strain="QM6a"
                  /db_xref="taxon:51453"
                  /clone="trico82xn19"
                  /dev_stage="mycelial"
                  /clone_lib="T.reesei mycelial culture, Version 6 October
                  2003"
                  /note="Vector: pREP3Y; Site 1: Not I/Sal I; Mycelial
                  culture grown from 24 hrs to 6 days with varying Carbon
                  and Nitrogen sources and concentrations."

```

ORIGIN

Alignment Scores:		
Pred. No.:	2,666-06	707
Score:	415.50	92
Percent Similarity:	55.51%	34
Best Local Similarity:	40.53%	90
Query Match:	8.09%	11
DB:	14	3
		Gaps:
		Length:
		Matches:
		Conservative:
		Mismatches:
		Indels:

US-09-917-376-1 (1-957) x CF880713 (1-707)

423	Qy	LysLeuGlyTrpMetAspGluAlaMetAlaIleAspPropheAsnSerAspArgMetLeu	442
		:::	
33	Db	CGCCTCGCGTGGATGTGAGTCTTCGAGATTGACCAACCGCAGCAACCATGGCTC	92
		:::	
443	Qy	TyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGly	462
		:::	
93	Db	TACGGCACCGGAATGACAATCTTTGGCGGCCACGATCTCACCACTGGGACACGCGCCAC	152
		:::	
463	Qy	GlnIleHisIleAlaProMetValIlyGlyLeuGluGluThrAlaValAsnAspLeulle	482
		:::	
153	Db	AATGTGTCAATCCAATCACTGGCAGACGGCATCGAGGAATCTCGCTCCAGCACTGGCC	212
		:::	
483	Qy	SerProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHis	502
		:::	
213	Db	TCTGCACCCGCGGAAGCAGCTATTGGCCGAGTCGCGAGACGACACGGCTTACCTTT	272
		:::	
503	Qy	Ala-----AspValThrAlaValProSerThrIlePheThrSerProValPheThr	519
		:::	
273	Db	GCCAGCAGAAACACCTTCGGGACATCGCCGAGACGGTCTGGGCAACGCCACATGGGCC	332
		:::	
520	Qy	ThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySer	539
		:::	
333	Db	ACCTCGACGAGCGTCGACTACGCCGGAGACTCGGTCAAGAGCGTCTCGGCTCGGCAAC	392
		:::	
540	Qy	PheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLys	559
		:::	
393	Db	ACGCCCGGCACGCACAG-----GTGGCCATCTCTGTCGACGGCGGCGCG	437
		:::	
560	Qy	AsnTrpPheGlnGlySerGluProGlyGlyValThrThrGlyGlyThrValAlaAlaSer	579
		:::	
438	Db	ACGTGGAGCATCGACTACCGCGGCCGACACGTCCTAAGAACCGCGGACGGTGCCTATT	497
		:::	
580	Qy	AlaAspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValTyrAla	599
		:::	
498	Db	GCCACCGCGCACACGATCCCTCTGTGTCGACCGCCCTCTGTCGGC-----GTG	548
		:::	
600	Qy	ValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArg	619
		:::	
549	Db	TCGCAGTTCCAGGCACCTTTCCTCTCGTCTCGAGCCTTGCCTCGCGGGCGCGTCA	608
		:::	
620	Qy	SerAspArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSer	639
		:::	
609	Db	TCGGACAAAGAACCAACAGCGTCTCTACGCCGCGCTCCGGATCGACCTTTTACGT	668
		:::	
640	Qy	ThrAspGlyGlyValThrPhe	646
		:::	
669	Db	AAGGACACCGCGCAGCAGCTTC	689
		:::	

RESULT 6	CB907625	LOCUS	tric082xn19	782 bp	linear	EST 02-JUL-2003
		DEFINITION	tric082xn19 T.reesei mycelial culture, Version 3 april Hypocrea jecorina cDNA clone tric082xn19, mRNA sequence.			

EST. 07025.1 GI:30122283
 KEYWORDS
 SOURCE
 ORGANISM
 Hypocrea jecorina (anamorph: Trichoderma reesei)
 Hypocrea jecorina
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreales; Hypocreaceae; Hypocremycetidae; 1 (bases 1 to 782)
 REFERENCE
 Foreman, P.K., Brown, D.E., Dankmeyer, L., Dean, R.

REFERENCE

1 Foreman, P.K.: 1978, 'The evolution of the Hymenoptera', pp. 1-60 in J. Mayr (ed.), *Evolutionary Biology*, University of Chicago Press, Chicago.

AUTHORS

Foreman, P.K., Brown, D.E., Dankmever, L., Dean, R., Diener, S.,
Hoffmann, B., Krombein, K.V., LaSalle, J., Lieke, W., Muesebeck,
J.C., Peterson, B.J., Polak, J.P., Schulze, H.P., Smith, A.B.,
Tromblay, L.F., Van Dine, J.M., Williams, G.A., and Wilson,
L.O.: 1982, 'A new classification of the Hymenoptera', *Hymenoptera*,
Nomenclature, pp. 1-100.

TITLE	Dunn-Coleman, N.S., Goedegebuur, F., Houfek, T.D., England, G.J., Kelley, A.S., Meerman, H.J., Mitchell, T., Mitchinson, C., Olivares, H.A., Teunissen, P.J., Yao, J., and Ward, M. Transcriptional regulation of biomass-degrading enzymes in the filamentous fungus <i>Trichoderma reesei</i>
JOURNAL	J. Biol. Chem. 278 (34), 31988-31997 (2003)
MEDLINE	22803314
PUBMED	12788920
COMMENT	Contact: Pamela K. Foreman Genencor Intl. 925 Page Mill Road, Palo Alto, CA 94304, USA Tel: (650) 846-7635 Fax: (650) 621-7817 Email: Pforeman@genencor.com Seq primer: LT-F1 primer.

	FEATURES	SOURCE
1.	100% Cotton	100%
2.	100% Polyester	100%
3.	100% Nylon	100%
4.	100% Rayon	100%
5.	100% Silk	100%
6.	100% Wool	100%
7.	100% Leather	100%
8.	100% Rubber	100%
9.	100% Metal	100%
10.	100% Glass	100%
11.	100% Plastic	100%
12.	100% Paper	100%
13.	100% Fabric	100%
14.	100% Wood	100%
15.	100% Stone	100%
16.	100% Concrete	100%
17.	100% Brick	100%
18.	100% Steel	100%
19.	100% Aluminum	100%
20.	100% Copper	100%
21.	100% Silver	100%
22.	100% Gold	100%
23.	100% Platinum	100%
24.	100% Titanium	100%
25.	100% Carbon Fiber	100%
26.	100% Kevlar	100%
27.	100% Fiberglass	100%
28.	100% Epoxy Resin	100%
29.	100% Polyurethane	100%
30.	100% Acrylic	100%
31.	100% Latex	100%
32.	100% Silicone	100%
33.	100% Neoprene	100%
34.	100% Spandex	100%
35.	100% Lycra	100%
36.	100% Elastane	100%
37.	100% Terylene	100%
38.	100% Dacron	100%
39.	100% Polyester	100%
40.	100% Cotton	100%
41.	100% Nylon	100%
42.	100% Rayon	100%
43.	100% Silk	100%
44.	100% Wool	100%
45.	100% Leather	100%
46.	100% Rubber	100%
47.	100% Metal	100%
48.	100% Glass	100%
49.	100% Plastic	100%
50.	100% Paper	100%
51.	100% Fabric	100%
52.	100% Wood	100%
53.	100% Stone	100%
54.	100% Concrete	100%
55.	100% Brick	100%
56.	100% Steel	100%
57.	100% Aluminum	100%
58.	100% Copper	100%
59.	100% Silver	100%
60.	100% Gold	100%
61.	100% Platinum	100%
62.	100% Titanium	100%
63.	100% Carbon Fiber	100%
64.	100% Kevlar	100%
65.	100% Fiberglass	100%
66.	100% Epoxy Resin	100%
67.	100% Polyurethane	100%
68.	100% Acrylic	100%
69.	100% Latex	100%
70.	100% Silicone	100%
71.	100% Neoprene	100%
72.	100% Spandex	100%
73.	100% Lycra	100%
74.	100% Elastane	100%
75.	100% Terylene	100%
76.	100% Dacron	100%
77.	100% Polyester	100%
78.	100% Cotton	100%
79.	100% Nylon	100%
80.	100% Rayon	100%
81.	100% Silk	100%
82.	100% Wool	100%
83.	100% Leather	100%
84.	100% Rubber	100%
85.	100% Metal	100%
86.	100% Glass	100%
87.	100% Plastic	100%
88.	100% Paper	100%
89.	100% Fabric	100%
90.	100% Wood	100%
91.	100% Stone	100%
92.	100% Concrete	100%
93.	100% Brick	100%
94.	100% Steel	100%
95.	100% Aluminum	100%
96.	100% Copper	100%
97.	100% Silver	100%
98.	100% Gold	100%
99.	100% Platinum	100%
100.	100% Titanium	100%

```

1. 0.782
source
/organism="Hypocrea jecorina"
/mol_type="mRNA"
/strain="QM6a"
/db_xref="taxon:51453"
/clone="tric082xm19"
/dev_tag="mycelia"

```

ORIGIN

Alignment Scores:		
Pred. No.:	2.95e-06	782
Score:	415.50	Matches:
Percent Similarity:	55.51%	Conservative:
Best Local Similarity:	40.53%	Mismatches:
Query Match:	8.09%	Indels:
DB:	14	Gaps:

US-09-917-376-1 (1-957) x CB907625 (1-782)

Qy	423	LysLeuGlyTrpMetAspGluAlaMetAlaIleAspPropheAsnSerAspArgMetLeu	442
Db	108	CGCTCCGGCTGGATGATTGATCTTCGAGATTGACCAACCGCAGCAACCACTGGCTC	167
Qy	443	TyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGly	462
Db	168	TACGGCACCGGAATGCAATCTTTGGCGGCCACGATCTCACCACTGGGACACCGGCCAC	227
Qy	463	GlnIleHisIleAlaProMetValLysGlyLeuGluGluThrAlaValAsnAspLeulle	482
Db	228	AATGTGTCAATCCAATCACTGGCAGACGGCATCGAGGAATTCTCGTCGACGACCTGGCC	287
Qy	483	SerProSerSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHis	502
Db	288	TCTGCACCCGCGGAAGCGAGCTATTATGGCCGAGTCGGAGACGACAAACGGCTTCACTTT	347
Qy	503	Ala-----AspValThrAlaValProSerThrIlePheThrSerProValPheThr	519
Db	348	GCCAGCAGAAACGACCTCGGGACATCGCCGACAGCGTCTGGGCAACGCCCACTGGGCC	407
Qy	520	ThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySer	539
Db	408	ACCTCGACGAGCGTCGACTACGCCGGNACTCGGTCAAGAGCGTCTCGCGTCGGCAAC	467
Qy	540	PheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLys	559
Db	468	ACGCCCGGCACCAACAG-----GTGGCCATCTCTGTCGACGCGCGCGCG	512
Qy	560	AsnTrpPheGlnGlySerGluProGlyValThrThrGlyGlyGlyThrValAlaAlaSer	579
Db	513	ACGTGGAGCATCGACTACGGCGCCGACACGTCCTCAAGACGGCGGACCGTGGCTATTTCG	572
Qy	580	AlaAspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValTyrAla	599

Pred. No.: 0.000243 Length: 748
 Score: 362.00 Matches: 72
 Percent Similarity: 83.91% Conservative: 1
 Best Local Similarity: 82.76% Mismatches: 14
 Query Match: 7.05% Indels: 0
 DB: 29 Gaps: 0

US-09-917-376-1 (1-957) x CE342004 (1-748)

Qy 782 GlyAlaProSerGlySerProSerProSerValSerProSerAlaSerProSerLeuSer 801
 Db 194 GGCCACCTTCACCTTCACCTTCCTTCACCTTCACCTTCACCTTCACCTTCCT 253
 Qy 802 ProSerProSerProSerProSerProSerProSerProSerProSerProSer 821
 Db 254 CCTTCACCTTCCTTCACCTTCACCTTCCTTCACCTTCACCTTCCTTCACCTTCA 313
 Qy 822 SerSerProSerProSerProSerProSerProSerProSerProSerProSer 841
 Db 314 CCTTCACCTTCCTTCACCTTCCTTCACCTTCCTTCACCTTCCTTCACCTTCA 373
 Qy 842 AlaSerProSerProSerProSerProSerProSerProSerProSerProSer 861
 Db 374 CCTTCCTTCACCTTCACCTTCCTTCACCTTCCTTCACCTTCCTTCACCTTCA 433
 Qy 862 ProThrProSerSerPro 868
 Db 434 CCTTCACCTTCACCTTCCTTCCT 454

RESULT 9

CF868882 751 bp mRNA linear EST 31-OCT-2003
 LOCUS trico16xm14.b1 T.reesei mycelial culture, Version 6 October 2003
 DEFINITION Hypocrea jecorina cDNA clone trico16xm14, mRNA sequence.

ACCESSION CF868882

VERSION

KEYWORDS

SOURCE CF868882.1 GI:38123564

ORGANISM Hypocrea jecorina (anamorph: Trichoderma reesei)

REFERENCE 1 (bases 1 to 751)

AUTHORS Diener,S.E., Dankmeyer,L., Dunn-Coleman,N., Houfek,T.D., Mitchell,T.K., van Solingen,P., Teunissen,P.J.M., Ward,M. and Dean,R.A.

TITLE Analysis of the protein processing and secretion pathways in a

JOURNAL Trichoderma reesei EST dataset

COMMENT Unpublished (2003)

Contact: Ralph A. Dean

Fungal Genomics Laboratory

North Carolina State University

Campus Box 7251, Raleigh, NC 27695, USA

Tel: 919-513-0020

Fax: 919-513-0024

Email: ralph.dean@ncsu.edu

Seq primer: LT-P1 primer.

Location/Qualifiers

1..751

/organism="Hypocrea jecorina"

/mol_type="mRNA"

/strain="QM6a"

/db_xref="taxon:51453"

/clone="trico16xm14"

/dev_stage="mycelia"

/clone_lib="T.reesei mycelial culture, Version 6 October 2003"

/note="Vector: pREP3Y; Site_1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."

FEATURES

source

1..751

/organism="Hypocrea jecorina"

/mol_type="mRNA"

/strain="QM6a"

/db_xref="taxon:51453"

/clone="trico16xm14"

/dev_stage="mycelia"

/clone_lib="T.reesei mycelial culture, Version 6 October 2003"

/note="Vector: pREP3Y; Site_1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."

ORIGIN

Alignment Scores:

Pred. No.: 0.000255 Length: 751

Score: 361.50 Matches: 88
 Percent Similarity: 51.61% Conservative: 40
 Best Local Similarity: 35.48% Mismatches: 103
 Query Match: 7.04% Indels: 17
 DB: 14 Gaps: 6

US-09-917-376-1 (1-957) x CF868882 (1-751)

Qy 453 AsnAspLeuThrValTrpAspSerGlyGlyGlnIleHisIleAlaProMetVallySGly 472
 Db 15 CACGATTCACCACTGGGACACGGCCACATGTGTCAATCACTCACTGGCAGCGG 74
 Qy 473 LeuGluThrAlaValAsnAspLeuLeuSerProSerGlyAlaProLeuLeuSer 492
 Db 75 ATCGAGGAATTCCTCGTCCAGGACCTGGCTCTGCACCGGGGGAAGCAGCTATTGGCC 134
 Qy 493 AlaLeuGlyAspLeuGlyGlyPheThrHisAla-----AspValThrAlaValPro 509
 Db 135 GCAGTCGGAGACGACCAACGGCTTCACCTTTCACGACGAAACACCTTCGGGACATCGCCG 194
 Qy 510 SerThrIlePheThrSerProValPheThrGlyThrSerValAspTrpAlaGluLeu 529
 Db 195 CAGACGGTCTGGGACACGCCACATGGGCCACTCGACGAGGTGCATACCGCGGGAAC 254
 Qy 530 AsnProSerIleIleValArgAlaGlySerPheAspProSerSerGlnProAsnAspArg 549
 Db 255 TCGGTCAAGAGCGTCTCGGTCGGCAACACCGCGCGGACGACAG-----302
 Qy 550 HisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGly 569
 Db 303 ---GTGGCCATCTCGTCCGACGGGGCGCGAGCTGGAGCATCGACTACCGCGCGGACACG 359
 Qy 570 ValThrThrGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaPro 589
 Db 360 TCCATGAACCGCGGACGCTGGCTATTTCGGCCGACGCGACACCATCTCTGGTGCACC 419
 Qy 590 GlyAspProGlyGlnProValValTrpAlaValGlyPheGlyAsnSerTrpAlaAlaSer 609
 Db 420 GCCTCGTCCGGC-----GTGACGCGCTCGAGTTCAGGGGAGCTTTGCTCCGTC 470
 Qy 610 GlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyr 629
 Db 471 TCGAGCCTCGCGCGGGCGCTCATCGCTCGGACAAAGACCAACACAGCGCTTCTTAC 530
 Qy 630 AlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProVal 649
 Db 531 GCGGCTCCGGATCGACCTTTTACGTACGAAGGACACCGCGGACGAGCTTC-----581
 Qy 650 AlaAlaGlyLeuProSerSerGlyAlaValGlyValMet-----PheHisAlaValPro 667
 Db 582 ACGCGCGG---CCAGCTGGGACGCGGAGGAGATCCGGGATATCGCTGCTACCCG 638
 Qy 668 GlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSerThrAsnGly 687
 Db 639 ACCACCGGGGACGTTGTATGTCTCGACCGGCGCGCATATTCGCTCCACAGACTCG 698
 Qy 688 GlySerSerTrpSerAlaIleThr 695
 Db 699 GGCACGACCTTGGCCAAAGTCTCC 722

RESULT 10

CB898982

LOCUS

DEFINITION

trico16xm14 T.reesei mycelial culture, Version 3 april Hypocrea

jecorina cDNA clone trico16xm14, mRNA sequence.

ACCESSION

CB898982

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

1 (bases 1 to 803)

CB898982 803 bp mRNA linear EST 02-JUL-2003
 LOCUS trico16xm14 T.reesei mycelial culture, Version 3 april Hypocrea
 DEFINITION jecorina cDNA clone trico16xm14, mRNA sequence.

ACCESSION

CB898982

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

1 (bases 1 to 803)

Hypocrea jecorina (anamorph: Trichoderma reesei)
 Hypocrea jecorina
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.

AUTHORS Foreman, P.K., Brown, D.B., Dankmeyer, L., Dean, R., Diener, S., Dunn-Coleman, N.S., Godegebuur, F., Houfek, T.D., England, G.J., Kelley, A.S., Meerman, H.J., Mitchell, T., Mitchinson, C., Olivares, H.A., Teunissen, P.J., Yao, J., and Ward, M.

TITLE Transcriptional regulation of biomass-degrading enzymes in the filamentous fungus *Trichoderma reesei*

JOURNAL J. Biol. Chem. 278 (34), 31988-31997 (2003)

MEDLINE 22803314

PUBMED 12788920

COMMENT Contact: Pamela K. Foreman
Genencor Intl.
925 Page Mill Road, Palo Alto, CA 94304, USA
Tel: (650) 846-7635
Fax: (650) 621-7817
Email: pforeman@genencor.com
Seq primer: LT-F1 primer.

FEATURES
source
1..803
Location/Qualifiers
organism="Hypocrea jecorina"
/mol_type="mRNA"
/strain="QM6a"
/db_xref="taxon:51453"
/clone="trio016xm14"
/dev_stage="mycelia"
/clone_lib="T. reesei mycelial culture, Version 3 april"
/note="vector: pREP4y; Site 1: Not 1/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."

ORIGIN

Alignment Scores:			
Pred. No.:	0.000273	Length:	803
Score:	361.50	Matches:	88
Percent Similarity:	51.61%	Conservative:	40
Best Local Similarity:	35.48%	Mismatches:	103
Query Match:	7.04%	Indels:	17
DB:	14	Gaps:	6

US-09-917-376-1 (1-957) x CB898982 (1-803)

Qy	453	AsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMetValLysGly	472
Db	67	CACGATTTCACCACTGGGACACGGCCACAAATGTGTCATCAATCACTCTGGCAGACGGC	126
Qy	473	LeuGluGluThrAlaValAsnAspLeuIleSerProProSerGlyAlaProLeuIleSer	492
Db	127	ATCGAGGAAATCTCGTTCAGACTGGCTCTGCACCCGGCGGAGGAGCTATTGGCC	186
Qy	493	AlaLeuGlyAspLeuGlyGlyPheThrHisAla-----AspValThrAlaValPro	509
Db	187	GCAGTCGGGAGACGACACACGGCTTCACCTTTGCCAGCAGAAACGACCTCGGGACATCGCGC	246
Qy	510	SerThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyrAlaGluLeu	529
Db	247	CAGACGGTCTGGGCAAGCCACATGGGCCACCTTCGACGAGCGTGCATCGCCGGGAAAC	306
Qy	530	AsnProSerIleIleValArgAlaGlySerPheAspProSerGlnProAsnAspArg	549
Db	307	TCGTCACAGAGGTGTCGGCGTGGCAACACCCCGCGCAGCAACAG-----354	
Qy	550	HisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGly	569
Db	355	---GTGGCCATCTGTCGACGGCGCGCAGCTGGAGCATCGACTACGGCGCGCACACG	411
Qy	570	ValThrThrGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaPro	589
Db	412	TCCATGAACGGCGGACGGTGGCCCTATTTCGGCCGACGGCGACACGATCCTCTCGTCGAC	471
Qy	590	GlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSer	609
Db	472	GCCTCGTCGGC-----GTGACGGCTCGCAGTTCAGGGCAGCTTTGCTCTCGTC	522
Qy	610	GlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyr	629

source

ORIGIN

Alignment Scores:

US-09-917-376-1 (1-957) x AZ374273 (1-259)

Search completed: May 14, 2004, 16:17:46
Job time : 5637 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 14, 2004, 08:53:04 ; Search time 158 Seconds
(without alignments)
3361.316 Million cell updates/sec

Title: US-09-917-376-1

Perfect score: 5135

Sequence: 1 MRSSENRILTMRSRLVSL.....RASFGSNPATPTDYLOX 957

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-Q=/cg2_1/USPIO_spool/US0917376/runat_14052004_085257_4644/app_query.fasta.1.1095
-DB=issued Patents NA -QFWT=fastap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blomsum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFWT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0917376 -CGN 1 1 75 -runat_14052004_085257_4644 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBIOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:*

1: /cg2_6/ptodata/2/ina/5A COMB.seq:*

2: /cg2_6/ptodata/2/ina/5B COMB.seq:*

3: /cg2_6/ptodata/2/ina/6A COMB.seq:*

4: /cg2_6/ptodata/2/ina/6B COMB.seq:*

5: /cg2_6/ptodata/2/ina/PCTUS COMB.seq:*

6: /cg2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	409	8.0	11707	3	US-09-136-574A-1
2	345	6.7	18596	3	US-09-318-448-11
3	345	6.7	18597	4	US-09-362-665-8
4	345	6.7	18597	4	US-09-363-333-8
5	334	6.5	6416	3	US-09-136-574A-2
6	319.5	6.2	8211	4	US-09-252-991A-13656
7	296.5	5.8	4767	4	US-09-410-551B-28
8	293	5.7	4818	4	US-09-410-551B-32
9	288.5	5.6	4131	4	US-09-252-991A-13773
10	288.5	5.6	4571	4	US-09-410-551B-18
11	288	5.6	288	4	US-09-119-507B-111
12	288	5.6	288	4	US-09-547-693-111

c 13	282.5	5.5	441529	3	US-09-103-840A-1	Sequence 1, Appli
c 14	281.5	5.5	4403765	3	US-09-103-840A-2	Sequence 2, Appli
c 15	280	5.5	3129	4	US-09-252-991A-13873	Sequence 13873, A
c 16	279	5.4	4547	4	US-09-410-551B-22	Sequence 22, Appl
c 17	273	5.3	4674	4	US-09-410-551B-26	Sequence 26, Appl
c 18	269	5.2	4478	4	US-09-410-551B-16	Sequence 16, Appl
c 19	266	5.2	4188	4	US-09-252-991A-13774	Sequence 13774, A
c 20	266	5.2	77536	4	US-09-410-551B-1	Sequence 1, Appli
c 21	265.5	5.2	77536	4	US-09-410-551B-1	Sequence 1, Appli
c 22	265	5.2	4737	4	US-09-410-551B-30	Sequence 30, Appl
c 23	260	5.1	3337	2	US-08-072-610-1	Sequence 1, Appli
c 24	260	5.1	3337	2	US-08-719-822B-1	Sequence 1, Appli
c 25	260	5.1	3337	3	US-09-092-458-1	Sequence 1, Appli
c 26	258	5.0	4725	4	US-09-410-551B-24	Sequence 24, Appl
c 27	257	5.0	4466	4	US-09-410-551B-20	Sequence 20, Appl
c 28	257	5.0	12001	1	US-08-458-568A-11	Sequence 11, Appl
c 29	255.5	5.0	3979	4	US-09-865-621A-4	Sequence 4, Appli
c 30	255.5	5.0	4810	4	US-09-865-621A-3	Sequence 3, Appli
c 31	255.5	5.0	5163	4	US-09-865-621A-7	Sequence 7, Appli
c 32	255.5	5.0	5509	4	US-09-865-621A-1	Sequence 1, Appli
c 33	251	4.9	4403765	3	US-09-103-840A-2	Sequence 2, Appli
c 34	250	4.9	441529	3	US-09-103-840A-1	Sequence 1, Appli
c 35	248	4.8	1926	4	US-09-249-585A-4	Sequence 4, Appli
c 36	248	4.8	1931	2	US-09-130-114-2	Sequence 2, Appli
c 37	247	4.8	3979	4	US-08-646-242-6	Sequence 6, Appli
c 38	238	4.6	47981	4	US-09-679-279-1	Sequence 1, Appli
c 39	237.5	4.6	25165	4	US-09-453-702B-39	Sequence 39, Appl
c 40	237	4.6	3427	4	US-09-865-621A-8	Sequence 8, Appli
c 41	237	4.6	3518	4	US-09-865-621A-5	Sequence 5, Appli
c 42	237	4.6	4647	4	US-09-252-991A-5730	Sequence 5730, Ap
c 43	234	4.6	6414	4	US-09-134-001C-1626	Sequence 1626, Ap
c 44	232	4.5	13842	3	US-09-105-537-30	Sequence 30, Appl
c 45	232	4.5	36778	3	US-09-105-537-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1

; Sequence 1, Application US/09136574A

; Patent No. 6294366

; GENERAL INFORMATION:

; APPLICANT: Farrington, Graham K.

; Anderson, Paige

; Gibbs, Moreland

; Bergquist, Peter

; Daniels, Roy

; Morgan, Hugh W.

; Williams, Diane P.

; TITLE OF INVENTION: Compositions and Methods for

; Treating Cellulose Containing

; Cellulase Enzyme Compositions

; NUMBER OF SEQUENCES: 49

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Howson and Howson

; STREET: Spring House Corporate Center, P.O. Box 457

; CITY: Spring House

; STATE: PA

; COUNTRY: USA

; ZIP: 19477

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/136,574A

; FILING DATE: 19-Aug-1998

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/932,571

; FILING DATE: September 19, 1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Bak, Mary E.

Sequence 1, Appli

Sequence 2, Appli

Sequence 13873, A

Sequence 22, Appl

Sequence 26, Appl

Sequence 16, Appl

Sequence 13774, A

Sequence 1, Appli

Sequence 1, Appli

Sequence 30, Appl

Sequence 1, Appli

Sequence 1, Appli

Sequence 24, Appl

Sequence 20, Appl

Sequence 11, Appl

Sequence 4, Appli

Sequence 3, Appli

Sequence 7, Appli

Sequence 1, Appli

Sequence 2, Appli

Sequence 1, Appli

Sequence 4, Appli

Sequence 5, Appli

Sequence 5730, Ap

Sequence 1626, Ap

Sequence 30, Appl

Sequence 5, Appli

232	Asp	Ile	Gln	Gly	Val	Val	Trp	Val	Ala	Ala	Phe	Asp	Leu	Ser	Ser	Ser	Ser	Leu	Gly	Gln	Ala	251					
1197	GTA	AAT	AC	AC	ACA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	1241					
252	Ser	Leu	Thr	Ile	Phe	Val	Gly	Val	Ala	Ala	Asp	Pro	Asn	---	---	---	---	---	---	---	---	267					
1242	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	1289					
268	Trp	Ser	Arg	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	279					
1290	TTT	TCA	AGT	GGT	ATT	TCA	AA	AT	TGC	CA	TCC	GAG	GGT	TGG	CAG	CA	AA	GGG	CA	AGG	CGG	1349					
280	Pro	Gly	Ala	Pro	Thr	Gly	Phe	Ile	Pro	His	Leu	Pro	His	Leu	Gly	Val	Phe	Asp	Pro	Val	Asn	His	Val	299			
1350	ACA	GT	AAAA	CA	CAG	AT	AG	CGT	TTG	TGG	CAT	AT	AG	TAT	AG	---	---	---	---	---	---	---	1394				
300	Leu	Tyr	Ile	Ala	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	313				
1395	TTG	TAC	GT	CAG	TGG	GA	AG	AC	GT	CA	AA	AT	TGC	AT	G	GT	G	C	A	G	A	TTC	CGG	TAG	AT	1454	
314	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	328			
1455	TTG	GA	CAC	AGG	TAA	AGT	GT	ATA	AAA	TAA	AGT	TTT	GGT	TTT	AT	CAG	A	C	A	C	A	CA	CT	GGT	TCA	AA	1514
329	Thr	Arg	Ile	Ser	Pro	Val	Pro	Ser	Thr	Asp	Thr	Ala	Asn	Asp	---	---	---	---	---	---	---	---	---	---	347		
1515	CA	AAA	AAT	GT	CA	TAA	CT	AT	CA	AA	CA	AG	AT	TT	TG	C	A	C	A	T	C	C	T	C	CA	CA	1574


```
; Patent No. 6210950
; GENERAL INFORMATION:
; APPLICANT: Johnson, William G.
; APPLICANT: Stenroos, Edward S.
; TITLE OF INVENTION: METHODS FOR DIAGNOSING, PREVENTING, AND TREATING
; TITLE OF INVENTION: DEVELOPMENTAL DISORDERS
; FILE REFERENCE: 601-1-057
; CURRENT APPLICATION NUMBER: US/09/318,448
; CURRENT FILING DATE: 1999-05-25
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 18596
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-318-448-11

Alignment Scores:
Pred. No.: 7,74e-08 Length: 18596
Score: 345.00 Matches: 71
Percent Similarity: 83.91% Conservative: 2
Best Local Similarity: 81.61% Mismatches: 13
Query Match: 6.72% Indels: 1
DB: 3 Gaps: 0

US-09-917-376-1 (1-957) x US-09-318-448-11 (1-18596)
Qy 783 AlaProSerGlySerProSerProSerValSerProSerAlaSerProSerLeuSerPro 802
Db 10388 TCACCATCACCATTCTCCATCTCCATCACCATCACCATCACCATCACCATCACCACCA 10329
Qy 803 SerProSerProSerProSerProSerProSerProSerProSerProSerProSer 822
Db 10328 TCACCATCTCCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCACCA 10269
Qy 823 SerProSerProSerProSerProSerProSerProSerProSerProSerProSerAla 842
Db 10268 TCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCACCA 10209
Qy 843 SerProSerProSerProSerProSerProSerProSerProSerProSerProSerPr 862
Db 10208 TCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCACCC 10149
Qy 862 oThrProSerSerPro 868
Db 10148 ATCTCCATCACCATCACCACCA 10130

RESULT 3
US-09-962-665-8/c
; Sequence 8, Application US/09962665
; Patent No. 6537759
; GENERAL INFORMATION:
; APPLICANT: Stanton, Jr., Vincent P.
; TITLE OF INVENTION: POLYPOLYGLUTAMATE SYNTHETASE GENE SEQUENCE
; TITLE OF INVENTION: VARIANCES HAVING UTILITY IN DETERMINING THE
; TITLE OF INVENTION: TREATMENT OF DISEASE
; FILE REFERENCE: 11926-015004
; CURRENT APPLICATION NUMBER: US/09/962,665
; CURRENT FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: 09/658,659
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 09/596,033
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: 09/357,743
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: 09/357,024
; PRIOR FILING DATE: 1999-07-19
; PRIOR APPLICATION NUMBER: 60/093,484
; PRIOR FILING DATE: 1998-07-20
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 18597

; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; LOCATION: 701_13751
; OTHER INFORMATION: n = c or a
; NAME/KEY: misc feature
; LOCATION: 716_1293, 2401, 2429, 2618, 3083, 3125, 3635, 4256, 4898,
; LOCATION: 5062, 5167, 11069, 13298, 14479, 14730, 14796, 15344, 15450,
; LOCATION: 15503, 15590, 15840, 16149
; OTHER INFORMATION: n = a or g
; NAME/KEY: misc feature
; LOCATION: 732_1379, 1590, 2488, 3212, 5006, 11238, 11422, 11686,
; LOCATION: 12598, 13171, 13645, 13782, 13806, 13813, 14586, 14788,
; LOCATION: 15042, 15546, 15770
; OTHER INFORMATION: n = c or t
; NAME/KEY: misc feature
; LOCATION: 1322_1688
; OTHER INFORMATION: n = c or g
; NAME/KEY: misc feature
; LOCATION: 2594, 11293, 16199, 16203
; OTHER INFORMATION: n = g or t
; NAME/KEY: misc feature
; LOCATION: 3619
; OTHER INFORMATION: n = a or t
; NAME/KEY: misc feature
; LOCATION: 14547
; OTHER INFORMATION: nucleotide in position 14547 is t, or absent
US-09-962-665-8

Alignment Scores:
Pred. No.: 7,74e-08 Length: 18597
Score: 345.00 Matches: 71
Percent Similarity: 83.91% Conservative: 2
Best Local Similarity: 81.61% Mismatches: 13
Query Match: 6.72% Indels: 1
DB: 4 Gaps: 0

US-09-917-376-1 (1-957) x US-09-962-665-8 (1-18597)
Qy 783 AlaProSerGlySerProSerProSerValSerProSerAlaSerProSerLeuSerPro 802
Db 10388 TCACCATCACCATTCTCCATCTCCATCACCATCACCATCACCATCACCATCACCACCA 10329
Qy 803 SerProSerProSerProSerProSerProSerProSerProSerProSerProSer 822
Db 10328 TCACCATCTCCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCACCA 10269
Qy 823 SerProSerProSerProSerProSerProSerProSerProSerProSerProSerAla 842
Db 10268 TCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCACCA 10209
Qy 843 SerProSerProSerProSerProSerProSerProSerProSerProSerProSerPr 862
Db 10208 TCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCACCC 10149
Qy 862 oThrProSerSerPro 868
Db 10148 ATCTCCATCACCATCACCACCA 10130

RESULT 4
US-09-963-333-8/c
; Sequence 8, Application US/09963333
; Patent No. 6664062
; GENERAL INFORMATION:
; APPLICANT: Stanton, Jr., Vincent P.
; TITLE OF INVENTION: THYMIDINE SYNTHASE GENE SEQUENCE VARIANCES
; TITLE OF INVENTION: HAVING UTILITY IN DETERMINING THE TREATMENT
; TITLE OF INVENTION: OF DISEASE
; FILE REFERENCE: 11926-015002
; CURRENT APPLICATION NUMBER: US/09/963,333
; CURRENT FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: 09/658,659
```

PRIOR FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: 09/596,033
PRIOR FILING DATE: 2000-06-15
PRIOR APPLICATION NUMBER: 09/357,743
PRIOR FILING DATE: 1999-07-20
PRIOR APPLICATION NUMBER: 09/357,024
PRIOR FILING DATE: 1999-07-19
PRIOR APPLICATION NUMBER: 60/093,484
PRIOR FILING DATE: 1998-07-20
NUMBER OF SEQ ID NOS: 16
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 8
LENGTH: 18597
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: 701, 13751
NAME/KEY: misc feature
LOCATION: 716, 1293, 2401, 2429, 2618, 3083, 3125, 3635, 4256, 4898,
LOCATION: 5062, 5167, 11069, 13298, 14479, 14730, 14796, 15344, 15450,
LOCATION: 15503, 15590, 15840, 16149
OTHER INFORMATION: n = a or g
NAME/KEY: misc feature
LOCATION: 732, 1379, 1590, 2488, 3212, 5006, 11238, 11422, 11686,
LOCATION: 12598, 13171, 13645, 13782, 13806, 13813, 14586, 14788,
LOCATION: 15042, 15546, 15770
OTHER INFORMATION: n = c or t
NAME/KEY: misc feature
LOCATION: 1322, 1688
OTHER INFORMATION: n = c or g
NAME/KEY: misc feature
LOCATION: 2594, 11293, 16199, 16203
OTHER INFORMATION: n = g or t
NAME/KEY: misc feature
LOCATION: 3619
OTHER INFORMATION: n = a or t
NAME/KEY: misc feature
LOCATION: 14547
OTHER INFORMATION: nucleotide in position 14547 is t, or absent
US-09-963-333-8

Alignment Scores:
Pred. No.: 7.74e-08 Length: 18597
Score: 345.00 Matches: 71
Percent Similarity: 83.91% Conservative: 2
Best Local Similarity: 81.61% Mismatches: 13
Query Match: 6.72% Indels: 1
DB: 4 Gaps: 0

US-09-917-376-1 (1-957) x US-09-963-333-8 (1-18597)

Qy 783 AlaProSerGlySerProSerProSerValSerProSerAlaSerProSerLeuSerPro 802
Db 10388 TCACCATCACCATTCTCCATCACCATCACCATCACCATCACCATCACCATCACC 10329
Qy 803 SerProSerProSerProSerProSerProSerProSerProSerProSerProSer 822
Db 10328 TCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATC 10269
Qy 823 SerProSerProSerProSerProSerProSerProSerProSerProSerProSerAla 842
Db 10268 TCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATC 10209
Qy 843 SerProSerProSerProSerProSerProSerProSerProSerProSerProSerPr 862
Db 10208 TCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATC 10149
Qy 862 oThrProSerSerProSerProSerProSerProSerProSerProSerProSerProSer 868
Db 10148 ATTCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATC 10130

RESULT 5
US-09-136-574A-2
Sequence 2, Application US/09136574A
Patent No. 6294366
GENERAL INFORMATION:
APPLICANT: Farrington, Graham K.
Anderson, Paige
Gibbs, Moreland
Bergquist, Peter
Daniels, Roy
Morgan, Hugh W.
Williams, Diane P.
TITLE OF INVENTION: Compositions and Methods for
Treating Cellulose Containing
Cellulase Enzyme Compositions
NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Center, P.O. Box 457
CITY: Spring House
STATE: PA
COUNTRY: USA
ZIP: 19477
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/136,574A
FILING DATE: 19-Aug-1998
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/932,571
FILING DATE: September 19, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: 1997US001/CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 6416 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-136-574A-2

Alignment Scores:
Pred. No.: 8.13e-08 Length: 6416
Score: 334.00 Matches: 219
Percent Similarity: 33.27% Conservative: 125
Best Local Similarity: 21.18% Mismatches: 314
Query Match: 6.50% Indels: 376
DB: 3 Gaps: 56

US-09-917-376-1 (1-957) x US-09-136-574A-2 (1-6416)

Qy 18 SerLeuLeuAlaAlaThrAlaSerPheAlaValAlaAlaLeuGlyValLeuProfile 37
Db 1189 TCTGTAGTGGCAGAAACA-----GCTGCATCATCTTGGGTGGCTTCAGTT 1233
Qy 38 AlaIleThrAlaSerProAlaHisAlaAla----- 47
Db 1234 GTATAAAGGAAGAAATTCACGAAGCAGCTTTATCTCCAAATGCCAAGACCTG 1293
Qy 48 -----ThrGlnPro-----TyrThrTrpSerAsnValAlaIle 59
Db 1294 TTTGAATTTGCCGATACCAAGAGAGTGGGGGTATCTGCTGCAACAGGTTCTTAC 1353

Qy 60 GlyGlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeuTyr 79
Db 1354 ACATCGGGTGGTATTATGATGACCTTGGATGG-----GCTGCTGATGGCTTAT 1404
Qy 80 ValArgThrAsp-----Ile 84
Db 1405 ATTCGACAAATGACAGTAGCTATTTCACGAAGCTGAAGAGTTGATGTCAGAAATATGCT 1464
Qy 85 GlyGlyMetTyrArgTrp-----AspAlaIleAsnGlyArgTrpIlePro 99
Db 1465 AATGGTACTAATACATGACACAAATGCTGGATGATGTTTCGATATGGAACATTTGATCATG 1524
Qy 100 LeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAla 119
Db 1525 CTTGCAAGATTACAGGGAAGAGTTA-----TATAAAGGAGCTGTGGAAAGAACTTA 1578
Qy 120 AspProIleAsnThrAsnLysValTrpAlaAlaValGlyMet---TyrThrAsnSerTrp 138
Db 1579 GACATTGGACTGACAGAATTACGTATACGCCGAAGGGATGGCATATCTGACAGGATGG 1638
Qy 139 AspProAsnAspGlyAlaIleLeuArgSerSerAspGlnGly-----152
Db 1639 -----GGTTCAATTAAGATATGCCACACAGCTGCATTTTTCAGCATGTGCTAT 1686
Qy 153 AlaThrTrpGln-----IleThr 158
Db 1687 GCAGACTGGTCAGGGTCGATTCGAACAAAGAACCAAAATATTGAACTTTGCAAAAGC 1746
Qy 159 ProLeuProPheLysLeuGlyGlyAsnMetProGlyArg-----GlyMetGly 174
Db 1747 CAGATTGACTGATGACTGGGTTCACAC-----CGTAGAAGTTTGTAGTAGGATTTGGC 1800
Qy 175 GluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGly 194
Db 1801 ACCAATTATCCACAAACATCGCATCAGCAAT-----1833
Qy 195 LysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPro 214
Db 1834 -----GCGCATAGTTTCATGGGTAAACGATGAAATACCA 1869
Qy 215 Asp-----ValGlyThrTyrIleAlaAsnProThrAspThrThrGly 228
Db 1870 GAGTATCACACACACATATTATATGAGCACTGGTTGGTGGTCTGCTGATGATGATGAT 1929
Qy 229 TyrGlnSerAspIleGlnValValTrp-----ValAlaPheAspLysSerSer 246
Db 1930 TATAATGATGACATTACCGATTATGTAACAAATGAGGTTCCTGCGGATTAATGCTGGA 1989
Qy 247 SerLeuGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProVal 266
Db 1990 ATTGTTGGTGCACCTGGCAAG-----2010
Qy 267 PheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPhe 286
Db 2011 -----ANGTACCAGTTATATGAGGTGAACCTATTGAT---2043
Qy 287 IleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThr-----304
Db 2044 -----GATTTTAAGCAATTGAACACCCACAAATGATGAAATTTTGTGTAATCAAAATTT 2100
Qy 305 SerAsnThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThr 324
Db 2101 GCGGAATTCACAGGTCCA-----AATTATACCGAAGTAATTTCTATATCTATAAT 2151
Qy 325 SerGlyThrTrpArgIleSerProValProSerThrAspThrAlaAsn---AspTyr 343
Db 2152 CGAACAGGATGG-----CCACCAAGGGTAACTGATAAATGATTTTAAATAT 2199
Qy 344 Phe-----GlyTyrSerGlyLeuThrIleAspArgGln 354
Db 2200 TTTATAGACCTAACCGAATTAATCCAGGAGGTATTTCG-----2238
Qy 355 HisProAsnThrIleMetValAlaThrGlnIleSerTrpProAspThrIleIlePhe 374

Db 2239 ----CCTGATTTGTCAAAGTTGACACA-----TAC 2265
Qy 375 ArgSerThrAspGlyGlyAlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsn 394
Db 2266 TACATCGAAGGAGGTAAATTAGCGTCTTACGATGGAC-----AAAAAT 2313
Qy 395 ArgSerLeuArgTyrValLeu---AspIleSerAlaGluProTrpLeuThrPheGlyVal 413
Db 2314 AGGAATATACATATGTTCTTGTGGATTTTAGTGAACCAAG-----ATA 2358
Qy 414 GluProAsnProProValProSerProLysLeuGlyTrp-----MetAspGlu 429
Db 2359 TATCTCGCGGTGAAGTTGAACACAAAGAGCGCTCAATTTAAATATCTGTTCCGCAG 2418
Qy 430 AlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeu 449
Db 2419 GGGTATCCATGGGATCTTACCNAAT---GATCCTTCATATAAGGATTAACCAAGTCAATTA 2475
Qy 450 TyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMet 469
Db 2476 GMAAAGAATAAATATATTGCGCATATGATAATAATAAT-----CTG 2517
Qy 470 ValLysGlyLeuGluThrAlaValAsnAspLeuIleSerPro-----484
Db 2518 GTATGGGGTTTAGAGCGCGGTGGCGCAACATCCACACCTGCACCAACATCAACCAACA 2577
Qy 485 -----ProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHis 502
Db 2578 CCAACCCCGACCCCAACACACACAGTACAGCA-----2610
Qy 503 AlaAspValThrAlaValProSerThrIlePheThr-----SerProValPheThr 519
Db 2611 -----ACGCGCAGCGCGACTCTTACACCGACACCGCGGGTCACT-2655
Qy 520 ThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySer 539
Db 2656 ACGGGAAGTGGTGTGAAGGTACTGTACAAGAACCAATGAGCAAGTCGCGAGCACAGTTCT 2715
Qy 540 PheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLys 559
Db 2716 ATAAAGCCG-----2724
Qy 560 AsnTrpPheGlnGlySerGluProGlyValThrThrGlyGlyThrValAlaAlaSer 579
Db 2725 ----TGGTTTAAG-----ATAGTAATGGAGG-----AGCAGCAGT 2757
Qy 580 AlaAspGlySerArgPhe-----ValTrp-----AlaProGlyAspProGlyGln 594
Db 2758 GTTGATCTTAGCAGGGTTAAGATAAGATACTGTGTACACAGTGGATGGTGAAGCCACAG 2817
Qy 595 ProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAla 614
Db 2818 AGTCGCTATGT-----GACTGG-----2835
Qy 615 AsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyrAlaLeuSerAsnGly 634
Db 2836 ----GCACAGATAGGGCAAGCAATGTGACA---TTCAATTTTGTCAAGCTTAGCAGCGGA 2889
Qy 635 ThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProValAlaAlaGlyLeuPro 654
Db 2890 GTG-----AGTGGAGCGGATTATTAC---CTGGAGGTAGGATTT---2925
Qy 655 SerSerGlyAlaValGlyValMetPheHisAlaValProGlyLysGlu---GlyAspLeu 673
Db 2926 AGCAGTGGAGCTGGCGAGTTGCGAG-----CCTGGTAAGGACACACAGGGGATATA 2973
Qy 674 TrpLeuAlaAlaSerSerGlyLeuTyrHisSerThrAsnGlyGlySerSerTrpSerAla 693
Db 2974 CAGTAGGTTTAAACAAGATGACTGGAGCAATTTACATACAGGACAGCAGTGTGTCATGG 3033
Qy 694 IleThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGlySerSer 713
Db 713 -----


```

Db 3034 TTCAGAGCATGACGAAT-----TATGGAGAGAAATCGCAGGTGACGCTG 3078
Qy 714 TyrProAlaValPheValValGlyThrIleGlyValThrGlyValAlaTyrArgSerAsp 733
Db 3079 TAT-----GTAGATGGT-----3090
Qy 734 AspCysGlyThrTrpValLeuIleAsnAspGlnHisGlnTyrGlyAsnTrpGly 753
Db 3091 -----GTTCTGGTA-----TGGGG 3105
Qy 754 GlnAlaIleThrClyAspHisAlaAsnLeuArgValTyrIleGlyThrAsnGlyArg 773
Db 3106 CAG-----3108
Qy 774 GlyIleValTyrGlyAspIleGlyAlaProSerGlySerProSerValSer 793
Db 3109 -----GAGCGGGAGGAGCG-----3123
Qy 794 ProSerAlaSerProSerLeuSerProSerProSerProSerProSerProSer 813
Db 3124 -----ACACCTGTCACCGACAAGCAGCAGCAACACCAACG 3156
Qy 814 ProSerProSerSerSerProSerProSerProSerProSerProSerProSer 833
Db 3157 CCAACTCCGACGACAAACCACTACACCTACACCGACCGCCGACGACGACGTGAGT 3216
Qy 834 ProSerArgSerProSerProSerAlaSerProSerProSerProSerProSer 853
Db 3217 GCACCCCAACACCGCGCAGCGGATCACCGGTAGTGCGGAGTACTGCGACGCGGAGT 3276
Qy 854 SerSerProSerSerProSerProSerProSerProSerProSerValSerGlyVal 873
Db 3277 GAGAGT-----TACGGTGGCGCTG 3294
Qy 874 LysValGlnTyrIleAsnAsnAspSerAlaProGlyAspAsnGlnIleLysProGlyLeu 893
Db 3295 AAGGTATGTATCGCAATGGAATTTAAGCAGCGCCGACGAATGATTTGAATCCTAAGATA 3354
Qy 894 GlnValValAsnThrGlySerSerSerValAspLeuSerThrValThrValArgTyrTrp 913
Db 3355 AAGATAGAGAATTTGCGGACGACGCGGTAGATCTTAGCAGGGTGAAGGTAAAGTACTCG 3414
Qy 914 PheThrArgAspGlyGlySerThrLeuValTyrAsnCysAspTrpAlaAlaIleGly 933
Db 3415 TACACATAGATGGTGAGCGGACACAGAGTGTA-----3447
Qy 934 CysGlyAsnIleArgAlaSerPheGlySerValAsnProAla 947
Db 3448 -----AGTGAGCGAGCAGCATAAATCCTGCG 3474

```

RESULT 6

```

US-09-252-991A-13656
; Sequence 13656, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 13656
; LENGTH: 8211
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13656

```

Alignment Scores:

```

Pred. No.: 5,65e-07 Length: 8211
Score: 319.50 Matches: 290
Percent Similarity: 32.56% Conservative: 133
Best Local Similarity: 22.32% Mismatches: 449
Query Match: 6.22% Indels: 432
DB: 4 Gaps: 69

US-09-917-376-1 (1-957) x US-09-252-991A-13656 (1-8211)

Qy 18 SerLeuLeuAlaAlaThrAlaSerPheAlaValAlaAlaLeuGlyValLeuProIle 37
Db 3326 GCAATGGCAGCAGCCTCAGCGGCACTGCGGAAACCGGGCAGCAGCGTGATCTCACCAGC 3385
Qy 38 AlaIleThr-----AlaSerProAlaHisAlaAlaThrThrGlnPro-----51
Db 3386 GCAACGGCAATCCGATCGCCGAGGTCAACCGCGAGCGGCACTGGACCTACACCC 3445
Qy 52 -----TyrThrTrpSerAsnVal-----AlaIleGly 60
Db 3446 CGTCCAGCCGATCGCCAAACGGTACTGTGTTC-AACGTGTGTGGCGGAGACGCGCGGT 3504
Qy 61 GlyGlyGlyPheValAspGlyIleValPheAsnGlyAlaProGlyIleLeuVal 80
Db 3505 AACAGCAGCCGCGCGGACCGTGCATTCAGCGCGCGCGCGCGCGCGGTGATC 3564
Qy 81 ArgThrAspIleGlyGlyMetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeu 100
Db 3565 AACCGGAGCAACGGCGTGTCTCATCAGCGGCACCGCGGAGCGCGGTGCCACCGTACCCTC 3624
Qy 101 LeuAspTrpValGlyTrpAsn-----AsnTrp 109
Db 3625 ACCGAT---GCCGCGGCAACCCGATAGGCGAGGTCAACCGCGGACGCGGCACTCG 3681
Qy 110 GlyTyr-----AsnGly-----ValValSerIleAlaAlaAsp 120
Db 3682 AGCTTTCAGCGCGGACGCGCGCGCCAAACGCGCGGTGATCGTCCACGCGCACCGAC 3741
Qy 121 Profile---AsnThrAsnLysValTrpAla-----AlaValGlyMetTyrThr 135
Db 3742 CCGACCGGCAATACCGCGCGCGCGCGCGCGCGCGCGCGGTGGAGCGGTGCCCGCGCGCG 3801
Qy 136 AsnSerTrpAspProAsnAspGlyAlaIleLeu-----146
Db 3802 CCGGTGATCGATCGAGCAACGGCACGACCATCAGCGCGACCGCGGAGCGCGGCGCAAG 3861
Qy 147 -----ArgSerSerAspGlnGlyAla 153
Db 3862 GTGATCTCACCAGCGCAACCGCAACCCGATCGCGGCAACCCACCGCGGACGCGGCGC 3921
Qy 154 ThrTrpGlnIleThrPro---LeuProPheLysLeuGly-----165
Db 3922 AACTGGACCTTCAGCGCGCGCGCGCGCGCGCGCGCGCGGTGGTGTCAACCGCGTGGCC 3981
Qy 166 -----GlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspPro 181
Db 3982 CAGGACCTCGGGCAATACCGCGCGCGCGCGCGCGCGCGCGCGGTGGTGGCGCG 4041
Qy 182 Asn-----AsnAspAsnIleLeuTyrPheGlyAlaProSerGly 194
Db 4042 AACACGCTGTGGTCAATCCGAGCAACGGCAACCTGTCTCAACGGTACCGCGGAGCGCGG 4101
Qy 195 LysGlyLeu-----TrpArgSerThrAspSer 203
Db 4102 AGCACCGGTGACCTTTCAGCGCGCGCAACCGGCAACCGGATCGCGCGGACGCGCGGTG 4161
Qy 204 GlyAlaThrTrp-----SerGlnMetThrAsnPheProAspValGlyThr 218
Db 4162 AGCGGCACTGGAGCTTCACCGCGCGGTCTCGCACTACCCAAAC-----GGCACC 4209
Qy 219 TyrIle---AlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyVal 237
Db 4210 GTGGTCAACGTGACCGCGGAGCGCGCGCGCGGCAATACCGAC-----GCTCCCGCT 4260

```

QY 238 TrpValAlaPheAspLysSerSerSerLeuGlyGln-----AlaSerLysThr 254
Db 4261 ACCACGACGGTGTCTTCCTCGCTGCGTGCATCCCGACGACGACGAGTTCG 4320
QY 255 IlePheValGlyValAlaAspProAsnAsnProValPhe----- 267
Db 4321 GTGATCAGCGGCACCGCGACCGCGGCAACACCATCATCATCATCCGATGGCAACGGCAAC 4380
QY 268 -----TrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaPro 283
Db 4381 CCGATTGGCCAGTCCACCCCGACGCGGAGCGGTAACTGGTCTTCACTCCAGGCATCCG 4440
QY 283 ----- 283
Db 4441 CTGCGGATGGCACGGTGTCAACGTGGTGGCGCGACGCCCAAGCAATGTCGACAGTCG 4500
QY 284 -----ThrGlyPheIleProHisIleGlyValPheAspProValAsn 297
Db 4501 CCGGGGTGATCACTGTGATGCGTGGCGCCCGCGCGCGCGGTGATCGATCCGAGCAAC 4560
QY 298 HisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyr-----AspGlySer 314
Db 4561 GGCACGAGTAAGCGGTACCGGAGGCGCGCGCGCGGTGATCTCACCAGTGGCGG 4620
QY 315 SerGlyAspValTrpLysPheSerValThr---SerGlyThrTrpThrArgIleSerPro 333
Db 4621 GGCACCCGATGGCGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4680
QY 334 ValPro-----SerThrAspThrAlaAsnAspTyr 343
Db 4681 ACCCGCTGGCCAAACCGCGACCGGTGATCAACCGCGTGGCCCGACCGCGCGCGCAATACC 4740
QY 344 PheGlyTyrSerGlyLeuThrIleAspArgGlnHisPro----- 356
Db 4741 AGCGGTCCGCGCAGCGTCACTCGTGCATCGCCCGCGCGCGCGCGGTGATCAATCCG 4800
QY 357 ---AsnThrIleMetValAlaThrGlnIleSerTrpTrpProAspThrIleIlePheArg 375
Db 4801 AGCAACGGAGTGTGTATCAGCGGTACCGCGGAAGCGCGGCGCGCGGTGATCTCTC----- 4854
QY 376 SerThrAspGlyGly-----AlaThrTrpThrArgIleTrpAspTrp 389
Db 4855 ---ACCGACGGCAACCGCAACCGCATCGGCCACCGGTCAACCGCGCGCGCGCAAGTGG 4911
QY 390 Thr-SerTyrProAsnArg-----SerLeuArgTyrValLeuAs 402
Db 4912 GCTTTCACCGCCCGCAGCGCGTGGCAATGGCAGCGGTGATCAATGCGCTGG----- 4963
QY 402 pIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnProProValProSerPr 422
Db 4964 ---CCAGGAGCGCGCGGCAACACAGCAGTCT-----CCACGCGGCA---CC 5007
QY 422 OlysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLe 442
Db 5008 GTGACTCGTGGCGCGCAGCA-GCCCGCGGTGATCGATCCGAGCAACCGGTAGC---GTGAT 5063
QY 442 uTyrGlyThr-----GlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAs 459
Db 5064 CGCGGTACCGCGAGGTGTGCCCGGTGATCTCACC----- 5105
QY 459 pSerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluThrAlaValAs 479
Db 5106 CGGCAACGGCAAC-----CCGATCGCGCAGGTCAACCGCGGTGATCGCGCGCA 5153
QY 479 nAspLeuIleSerProProSerGlyAlaProLeu-----IleSerAlaLe 494
Db 5154 CTGAGCTTCAAGCC-----GGCAGCGCGGTGTCCAATGGCAGCGGTGATCAATGCGGT 5207
QY 494 uGly---AspLeuGlyGlyPheThrHisAlaAspVal---ThrAlaValProSerThrI 512
Db 5208 GGCCCGAGGACGTGGCGGCAACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5267
QY 512 ePheThrSerProValPhe-----ThrThrGlyThrSerVal---AspTyrAlaGluLe 529

Db 5268 GCCGGCGCGCGGTGATCGACCGCAACCGCGACGCGTGTGCGCGGTACCGCGGAAGC 5327
QY 529 uAsnProSerIleIleValAlaGlySerPheAspProSerSerSerGlnProAsnAspAr 549
Db 5328 CGGTGCGACGGTGTCTCACCAGTGGCGGCGCAACCGCGATCGCGCGCGCGCGCG 5379
QY 549 gHisValAlaPheSerThrAspGlyGlyLysAsnTrp----- 561
Db 5380 ---ACCAGCGATGGCAGCGCAACTGGAGCTTCAACCGCGCGCGCGCGCGCGT 5426
QY 562 -----PheGlnGlySerGluProGlyGlyValThrThrGlyGly 574
Db 5427 GGCCACGGCACGGTGTGATCAATCGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCG 5486
QY 574 YThr----- 575
Db 5487 GACCAGCACCGGTGGACGGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5546
QY 576 -----ValAlaAlaSerAlaAspGlySerArgPheValTrpAlaProGlyAspProGl 593
Db 5547 CAGCGTGTGTCGGGTACCGCGGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5606
QY 593 yGlnProVal-----ValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGlnGl 611
Db 5607 CAACCCGATCGCGCAGGTCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5663
QY 611 yValPro---AlaAsnAlaGlnIleArgSer----- 620
Db 5664 CAGCGCGTGGCCAAACCGGTGATCAATCGGTGGCGCGCGCGCGCGCGCGCGCGCGCG 5723
QY 621 -----AspArgValAsnProLysThr---PheTyrAlaLe 631
Db 5724 CAGCG 5783
QY 631 uSerAsnGlyThrPheTyrArgSer-----ThrAs 641
Db 5784 GAGCAACCGGTACCGGTGATCAGCGGTACCGCGGAAGCGCGCGCGCGCGCGCGCGCG 5843
QY 641 pGlyGlyValThrPheGlnProValAlaAlaGlyLeu-ProSerSerGlyAlaValGlyV 661
Db 5844 CGGCGCG-----GGCAACCGGATACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5897
QY 661 alMetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyL 681
Db 5898 CTTCACTCCGGGCAAC-CCGCTGACCAACGCGCGGTGATCAATGCGGTGGCGCGCGCG 5956
QY 681 euTyrHisSerThrAsnGly-----GlySerSerTrpSerAlaIleThrGlyValSerS 699
Db 5957 CCGCGCGCAACACCGCGGTCCGCTCAGCACCGCGCGGTGCGCGCGCGCGCGCGCGCG 6016
QY 699 erAlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheV 719
Db 6017 CGGTGATCG 6076
QY 719 alValGlyThrIleGlyGly----- 725
Db 6077 TGATCTCAGCGATGGCAATGGTCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6136
QY 726 -----ValThrGlyAlaTyrArgS 732
Db 6137 ACTGGACCTTCAACCG 6196
QY 732 erAspAspCysGlyThrTrpValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnT 752
Db 6197 AGGACCG 6246
QY 752 rpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArg-ValTyrIleGlyThrAsn 771
Db 6247 GTGGCG 6306
QY 772 -----GlyArgGlyIleValTyrGlyAspIleGlyAlaProSerGlySer-ProSe 789

```
Db 6307 GAGTCGGCGCCAAAGTGTCTCACCAGCGGCAACCGCAACCGATCGCGGAGACCACC 6366
Qy 789 rProSerVal-----SerProSerAlaSer-----ProSerLeu----- 800
Db 6367 GCCAGCGGAGTGGCAACTGACCTTCAACCCCGGACCGCGCTGGCCAAACGGTACGGTG 6426
Qy 801 -----SerProSerPro-----SerProSerSerProSerProSerPro----- 816
Db 6427 ATCAACCGCGTTCGCGAAGACGCGCGGCAACCCAGCTCGCGGCAGCACCCAGCGGTG 6486
Qy 817 -----SerSerSerSerSerSerSerSerSerSerProSerPr 826
Db 6487 GACTCGGTGGCGCGTCCGCTCCGCTGCTGAGCATCAGCGCGGACCGCGGCTGCTGACC 6546
Qy 826 oSerPro-----SerProSerProSerProSerAr 836
Db 6547 GGCACCGCGGACCGGACAGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6606
Qy 836 gSerProSerProSerAlaSerProSerProSerSer-----Se 849
Db 6607 ATCAACCGTCACTCGTCAAGCGCGCGGCAACTTCAAGCTCGGCTGCGCGCGCGCTGATC 6666
Qy 849 rProSer-----ProSerSerSerProSerSerSe 859
Db 6667 ACCGCGCACTGATCGCGCGGTTGCGTGGACGCGCGCGCAACGTCAGCGGCGCGGCC 6726
Qy 859 rProSer-----ProThPr 864
Db 6727 ACCATCAACCGCGGACCTGGCGCGCGCGGACCATCAAGTGGCGGAGCGCGGATACC 6786
Qy 864 oSerSerSerProValSerGlyGlyVallyVallyVallyVallyVallyVallyVallyVal 884
Db 6787 TGGATCAACCGCGGAGATCGGCGGCGCGG-ATCCAGTGC----- 6825
Qy 884 oGlyAspAsnGlnIleLysProGlyLeuGlnValValAsnThrGlySerSerSerValas 904
Db 6826 ---GATGTCACGCGTCCGTCGACCATATGACGTCAGTGGC----- 6858
Qy 904 pLeuSerThrValThrValargTyrTrpPheThrArgAspGlyGlySerSerThrLeuVa 924
Db 6859 ---CAGTGGTCACTCGGTCAAGTTC---GCCGGCGAAGACGGCTACGAGCGCGAGGTGAC 6911
Qy 924 lTyrAsnCysAspTrpAlaAlaIleGlyCysGlyAsnIleArgAlaSerPhe----- 941
Db 6912 CCATACCTCAGCGCGCGGACATCGCGCGCGGCAACCTGACCTGACCTGACCTGACCTCC 6971
Qy 942 -GlySerValAsnPro-----AlaThrProThrAlaAsp 952
Db 6972 CGCGCGCATGGCGCGCTTCCCGGAGGCGCGCTCGACCGTCAACCGCGCAC 7020
```

RESULT 7

```
US-09-410-551B-28
; Sequence 28, Application US/09410551B
; Patent No. 6503737
; GENERAL INFORMATION:
; APPLICANT: KOSAN BIOSCIENCES, Inc.
; APPLICANT: REEVES, CHRISTOPHER
; APPLICANT: CHU, DANIEL
; APPLICANT: KHOSLA, CHAITAN
; APPLICANT: SANTI, DANIEL
; APPLICANT: WU, KAI
; TITLE OF INVENTION: POLYPEPTIDE SYNTHASE ENZYMES AND RECOMBINANT DNA
; FILE REFERENCE: 30062-20026.00
; CURRENT APPLICATION NUMBER: US/09/410,551B
; CURRENT FILING DATE: 1999-10-01
; PRIOR APPLICATION NUMBER: US 60/139,650
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/123,810
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 60/102,748
; PRIOR FILING DATE: 1998-10-02
; NUMBER OF SEQ ID NOS: 72
```

```
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 4767
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: DNA encoding synthetic
; OTHER INFORMATION: PKs synthase fragment
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3)..(4766)
US-09-410-551B-28
```

Alignment Scores:

```
Pred. No.: 4.23e-06 Length: 4767
Score: 296.50 Matches: 233
Percent Similarity: 35.16% Conservative: 101
Best Local Similarity: 24.53% Mismatches: 397
Query Match: 5.77% Indels: 223
DB: 4 Gaps: 46
```

US-09-917-376-1 (1-957) x US-09-410-551B-28 (1-4767)

```
Qy 9 LeuThrMetArgSerArgArgLeuValSerLeuLeuAlaAlaThrAlaSerPheAlaVal 28
Db 1827 CTGACCGAACAAGACCGGCTGCGGCTACCTGCGGCGTGCCTCGCGGCGGCGGATATA 1886
Qy 29 AlaAlaLeuGlyValLeuProIleAlaThrAlaSerProAlaHisAlaAlaThr 48
Db 1887 CGGCTGTGGATCGACGCTG-----GCGGTGACACGGTTCGGT-GTTCGAGCAGCCGCGC 1939
Qy 49 ThrGlnProTyr----- 52
Db 1940 GGTACTCTTGAGATGACACCGTCAACCGGACCGCGGTGACCGACCCCGAGATCGTGT 1999
Qy 53 -----ThrTrpSerAsnValAlaIleGlyGlyGly-----PheVal 65
Db 2000 TGTCTTCCCGGCGAGGGTGGCAGTGGCTGGGATGGCAGTGCACTCGCGGATTCGTC 2059
Qy 66 AspGlyIleValPheAsnGluGly-----AlaProGlyIleLeuTyrValArgThrAsp 83
Db 2060 GGTGGTGTTCGCGAGCGGATGGCGGAGTGTGCGGCGGCTTGGCGAGATTCGT---GGA 2116
Qy 84 IleGlyGlyMetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrp 103
Db 2117 CTGGGATCTGTTCAGGTTCTGGATGATCCGCGCGGTGG-TGGACCGGGTTGATG---TGG 2172
Qy 104 ValGlyTrpAsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsn 123
Db 2173 TCCAGCCC-GCTTCTGGCGGATG-----ATGGTTTCTTCTGGCGCGG----- 2213
Qy 124 ThrAsnLysValTrpAlaAlaValGlyMetTyrThrAsnSer---TrpAspProAsnAsp 142
Db 2214 -----GTGTGGCAGCGCGCGGTGTCGCGCGGATCGGTGATCGGCATTCGCGAG 2264
Qy 143 GlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGln----- 156
Db 2265 GGTGAGATCGCCGAGCTTGTGTGGCGGGTGGGTGTCACGTACGCGATGCCCGCGGATC 2324
Qy 157 IleThrProLeuProPheLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArg 176
Db 2325 GTGACCTTGGCAGCCAGCGCATCGCCCGGGCGCTGGCGGCGCGGCGGATGCATCC 2384
Qy 177 LeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGly 196
Db 2385 GTCGCCCTG---CCCGCGCAGGATGTCGAGTGGTGGCAGCGGCGC----- 2426
Qy 197 LeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspVal 216
Db 2427 ---TGGATCGCGCGCCCAACGCGCGCGCTCCACCGGTGATCGCGGCGCACCCCGAAGCG 2483
Qy 217 GlyThrTyrIleAlaAsnProThrAspThrThrGly----- 228
```

2484 Db GTGACCATGCTCTCACCCTCATGAGGACACAGGGGTGCGGTGCGCGGATCACCGTC 2543
229 Qy ---TyrGlnSerAspIleGlnGlyValValTrpVal-----AlaPheAspLysSer 244
2544 Db GACTATGCTCTGCACACCCCGCAGCTGAGCTGATCGCGACGAACTACTCGACATCACT 2603
245 Qy SerSerSerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsn 264
2604 Db AGGCACAGCAGCTCGCAGACCCCGCTCGTCCGCTGCTGCGACCGGTGCGACCGGCACCC--- 2660
265 Qy ProValPheTrp-----SerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAla 282
2661 Db -----TGGTGTGACACCCCGCTGAGCGGAGTACTGACCGAACCTCGGTGAA 2711
283 Qy ProThrGlyPheIleProHisLysGlyValPheAspProValAsnHisValLeuTyrIle 302
2712 Db CCGTGGTGTTCACCCCGCTGAGCGGAGTACTGACCGAACCTCGGTGTC 2771
303 Qy AlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSer 322
2772 Db GAGTCAAGCCAGCCCGCTGTTGTCAGCGCATGACGACGATGTCACCGTTGCC 2831
323 Qy Val-----ThrSerGlyThrTrpThrArgIleSerProValProSerThrAspThr 339
2832 Db ACGTGTGCTGACGACGCGCAGCGCAGCCCGCATG-----CTCACCGCCCTG 2879
340 Qy AlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIle 359
2880 Db GCACAGCCTAT-----GTCCAGCGGTACCGTGCAC----- 2912
360 Qy MetValAlaThrGlnIleSerTrpTrpProAspThrIleIlePheArgSerThrAspGly 379
2913 Db -----TGCCCGCCATCCTC----- 2927
380 Qy GlyAlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyr 399
2928 Db GGCACCAACACACCCCGGTACTGAGCTTCCGACCTTCCGACCTTCCACACACCGGTAC 2987
400 Qy ValLeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnProVal 419
2988 Db -----TGCTC-----GAGTGGCTCCCGCGCC 3011
420 Qy ProSerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAsp 439
3012 Db ACGGCC-----GACTCGGC 3026
440 Qy ArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAsp 459
3027 Db CACCCCGCTCTCGCACCGGAGTCCGCTCGCGG----- 3062
460 Qy SerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluThrAla---Val 478
3063 Db TCGCGCGCGGGTTCACGGTTCCTGCGCGCGCGTGGACCGCGGTGTTATC 3122
479 Qy AsnAsnLeuIleSerProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGly 498
3123 Db GCCGAACCTGGCTCGCGCGCGCGCGCCAC-----GACTCGCC 3164
499 Qy GlyPheThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProValPhe 518
3165 Db ACGGTCAACAGCTCGACGTCACCTCGCTCGCGCGGATCCGCGCGGAGGCC--- 3221
519 Qy ThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGly 538
3222 Db ACCCGCAGACCTGGGTGAT----- 3242
539 Qy SerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGly 558
3243 Db -----GAACCGCGCGCGCGCGCGGTTCACCGTCCACACCGCGGTGCGGAC 3296
559 Qy LysAsnTrpPheGlnGlySerGluProGlyGlyValThrThrGlyThrValAlaAla 578
3297 Db GCCCGGTGGAGCTGCACCGCGAG-----GGGTTCCTCGCGCGCGCGGTG---CCC 3347

579 Qy SerAlaAspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValTyr 598
3348 Db CAGCCCGAAGCGGTGACACCGCTGCGCCCG-----CGGGC-----CGGTGCC 3395
599 Qy AlaValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIle 618
3396 Db CGGACCGGTGCGCGCGGTGGCGCGCGGACGAGTCTTCTGTAAGCCGGAAGTC 3455
619 Qy ArgSerAspArgValAsnProLysThrPheTyrAla-----LeuSerAsnGlyThr 635
3456 Db GACAGC-----CCTGACGGCTTGTGGGCACACCCCGACCTGTCTGACGCGGTC 3503
636 Qy PheTyrArgSerThrAspGlyGlyValThrPheGlnProValAlaAlaGlyLeuProSer 655
3504 Db TTCTCCGCGGTGCGCGCGG-----AGCGCCAGCGCGACCGGATG-CGCGGACCTCGC 3556
656 Qy SerGlyAlaValGlyValMetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeu 675
3557 Db GGTGACGCGGTGCGACGCCCGGTGCGCGCTGCTCACCCCGCGGACGAGTGTGT 3616
676 Qy AlaAlaSerSerGlyLeuTyrHisSerThrAsnGlyGlySerSerTrpSerAlaIleThr 695
3617 Db CTGGAGCTCGCGCTTCGACGGTGGCGGATGCGGT-----GCTCACCGC 3664
696 Qy GlyValSerSerAla-----ValAsnValGlyPheGlyLysSerAlaProGlySerSer 713
3665 Db GGAGTGGTGGAGTGGCGGAGTGGCGGCGGATCCGACGAGTCCGACGCTCT 3724
714 Qy TyrProAlaValPheValValGlyThrIleGlyGlyVal-ThrGlyAlaTyrArgSerAs 733
3725 Db GTTTCGGCTTGAAGTGGTGGCGGAGCCACCTACGACGCTGCC-----GA 3775
733 Qy PAspCysGlyThrTrpValLeuIleAsn-----AspAspGlnHisGlnTyr 749
3776 Db CGAGTGGCGGAGGTACACCTCATCCGCCACACACCCCGACGACCCCGACGCC 3835
749 Qy rGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgVal----- 766
3836 Db CACCAACCCCGACACACACCGACCGCACACCAACACCAACACCGGCTCTCACCGC 3895
767 Qy -----TyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIle----- 780
3896 Db CTCCACACACCTCATCACCAACACACACCTCATCGTCCACACCGACCGACCC 3955
781 Qy ---GlyGlyAlaProSerGly-SerProSerPro-----SerValSerProSerAlas 797
3956 Db CCGAGCGCGCGGTACCGGCTCACCGGCTCACCGCACCAACACCAACCGCGCGCAT 4015
797 Qy erProSerLeuSerPro-----SerProSerProSerSerSerProS 811
4016 Db CCACCTCATCGAAACCCACACCCCGACCGCTCCCGCTCACCAACTCACCACT 4075
811 Qy erProSerPro-----SerProSerSerSerProSerSerProS 825
4076 Db CCACCAACCCCGCTACGCTCACCAACACCGCTCCACCGCGCGCAT 4135
825 Qy erProSerProSerProSerProSerProSerProSerProSerProS 845
4136 Db CACCAACCCCGACACACCGACCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4195
845 Qy erProSerSerSer-----ProSerProSerSerSerSerSerSerP 860
4196 Db CGCATCTCTCATCGCGGCGGTCCGCGACCTTCGCGGCGATCTCTCGCGCGCGCGCG 4255
860 Qy roSerProThrProSerSerPro 868
4256 Db CCACCCCGACACCTACTCTCTCTCC-4281

RESULT 8

US-09-410-551B-32

; Sequence 32, Application US/09410551B

; Patent No. 6503737

GENERAL INFORMATION:
APPLICANT: KOSAN BIOSCIENCES, Inc.
APPLICANT: REEVES, CHRISTOPHER
APPLICANT: CHU, DANIEL
APPLICANT: KHOSLA, CHAITAN
APPLICANT: SANTI, DANIEL
APPLICANT: WU, KAI
TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA
TITLE OF INVENTION: CONSTRUCTS THEREFOR
FILE REFERENCE: 30062-20026.00
CURRENT APPLICATION NUMBER: US/09/410,551B
CURRENT FILING DATE: 1999-10-01
PRIOR APPLICATION NUMBER: US 60/139,650
PRIOR FILING DATE: 1999-06-17
PRIOR APPLICATION NUMBER: US 60/123,810
PRIOR FILING DATE: 1999-03-11
PRIOR APPLICATION NUMBER: US 60/102,748
PRIOR FILING DATE: 1998-10-02
NUMBER OF SEQ ID NOS: 72
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 32
LENGTH: 4818
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: DNA encoding synthetic
OTHER INFORMATION: PKS synthase fragment
FEATURE:
NAME/KEY: CDS
LOCATION: (3)..(4817)
US-09-410-551B-32

Alignment Scores:
Pred. No.: 6,39e-06 Length: 4818
Score: 293.00 Matches: 235
Percent Similarity: 33.07% Conservative: 97
Best Local Similarity: 23.41% Mismatches: 379
Query Match: 5.71% Indels: 295
DB: 4 Gaps: 48

US-09-917-376-1 (1-957) x US-09-410-551B-32 (1-4818)

Qy 1 MetAspArgSerGluAsnIleArgLeuThrMetArgSer----- 13
Db 1798 CTGACCGCTCCCGCGCGCGCGCGCGTCCAGCAGCGCGGAGACCTTCCGCTGCTGTGT 1857
Qy 14 ArgArgLeuValSerLeuLeuAlaAlaThrAlaSerPheAlaValAlaAlaLeuGly 33
Db 1858 CGGCGGTTCCCGGAGGCACTCGACGAGCAGATCGGCGCGCTGCGCG----- 1905
Qy 34 ValLeuProIleAlaIleThrAlaSerProAlaHisAlaAlaThrGlnProTyrThr 53
Db 1906 -----CCTATC-----TCGACACCGCGCGCGCGTCCGACCGCGCGCGTGGCGC 1950
Qy 54 -----TipSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsn 71
Db 1951 AGACATGG----- 1959
Qy 72 GluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyMetTyr----- 88
Db 1960 -----CCCGGC-----GTACGCATTCACCCCGCGCGTACTGCTCGGGG 2001
Qy 88 ----- 88
Db 2002 ACACCGTATCGCGCTCCCGCGCGCGGACGAGCGAGCACTGCTTCTGCTACTCCG 2061
Qy 89 -----ArgTtpAspAla 92
Db 2062 GTCAGGCGACCCAGCATCCCGCGATGGCGAGCAGCTAGCCGATTCGTGGTGTGTTC 2121
Qy 93 AlaAsnGly----- 95
Db 2122 CCGAGCGAGTGGCGGAGTGTGGCGCGCGTTCGCGAGTTCGTGGACTGGGATCTGTCA 2181

Qy 96 -----ArgTtpIleProLeuLeuAspTtpValGlyTrpAsnAsnTtp 109
Db 2182 CGGTTCTGATGATCCGCGGTGGTGGACCGGGTGTGTGTGTCCAGCCC-GCTTCTCG 2240
Qy 110 GlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTtpAla 129
Db 2241 GCGATG-----ATGGTTTCTCTGGCCGCG-----GTGTGGCAG 2273
Qy 130 AlaValGlyMetTyrThrAsnSer---TtpAspProAsnAspGlyAlaIleLeuArgSer 148
Db 2274 GCGCGCGGTGGCGCGCGATGCGGTGATCGGCCATTGCGAGGTGAGATCGCCGAGCT 2333
Qy 149 SerAspGlnGlyAlaThrTtpGln-----IleThrProLeuProPhe 162
Db 2334 TGTGTGGCGGTGGCGGTGCTGCTACTAGCGATGCGCGCGGATCGTACCTTGGCAGCCAG 2393
Qy 163 LysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsn 182
Db 2394 GCGATCGCGCGCGCGCTGGCGGCGCGCGATGCGATCGCGCTG-----CCGCG 2450
Qy 183 AsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTtpArgSerThrAsp 202
Db 2451 CAGGATGTCGAGCTGTCGACGGGCG-----TGGATCGCGCGCCAC 2492
Qy 203 SerGlyAlaThrTtpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsn 222
Db 2493 AACGGCGCGCTCCCGGTGATCGCGGCGACCCCGGAGCGGTGACCATGTCTCTCACC 2552
Qy 223 ProThrAspThrGly-----TyrGlnSerAspIle 233
Db 2553 GCTCATGAGGCACAAAGGGTGGCGGTGCGGGATGATCGCGTACCTATGCTCGCACACC 2612
Qy 234 GlnGlyValValTtpVal-----AlaPheAspLysSerSerSerLeuGlyGln 250
Db 2613 CCGCACGTGAGTGTATCCGCGACGAACTACTTCGACATCATCAGTACGACAGTGTGCGAG 2672
Qy 251 AlaSerLysThrIlePheValGlyValAlaAlaProAsnAsnProValPheTtp----- 268
Db 2673 ACCCGCTGTCGCGTGTGCTGTCGACCGTGGACGCGAC-----TGGTTCGAC 2720
Qy 269 SerArgAspGlyGlyAlaThrTtpGlnAlaValProGlyAlaProThrGlyPheIlePro 288
Db 2721 AGCCGCTGGAGCGGAGTACTGTGTACCGGAACCTGCGTGAACCGGTGTTTCCACCCC 2780
Qy 289 HisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGly 308
Db 2781 GCCGTACGCCAGTTGCGAGCGCGCGACCGGTGTTGTCGAGGTGACGCGCGACCCG 2840
Qy 309 GlyProTyrAspGlySerSerGlyAspValTtpLysPheSerVal-----ThrSer 325
Db 2841 GTGTTGTGGAGCGATGGACGACGATGCTGTCAGTTCGCGTTCGCGTGTGTCGACGAC 2900
Qy 326 GlyThrTtpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGly 345
Db 2901 GGCAGCGCCACCGGATG-----CTCACCGCGCTGCGACAGCCCTAT----- 2942
Qy 346 TyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIle 365
Db 2943 GTCCACGCGGTCCCGTTCGAC-----GGCACCACCAACACCCG 2963
Qy 366 SerTtpTtpProAspThrIlePheArgSerThrAspGlyGlyAlaThrTtpThrArg 385
Db 2964 -----TGGCCCGCATCTCTC-----GGCACCACCAACACCCG 2996
Qy 386 IleTtpAspTtpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAla 405
Db 2997 GTACTGACCTTCGACCTACGCTTCCACACGAGCGGTAC----- 3038
Qy 406 GluProTtpLeuThrPheGlyValGlnProAsnProProValProSerProLysLeuGly 425
Db 3039 -----TGGCTC-----GAGTGGCTGCTCCCGCGCGCGCG----- 3068

Qy 426 TrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThr 445
Db 3069 -----GACTCGGCGCCACCGCTCGGCGAC 3095
Qy 446 GlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHis 465
Db 3096 GGAGTCGCGTCCCGCGG-----TCGCGGCGCGGCTTC 3131
Qy 466 IleAlaProMetValLysGlyLeuGluThrAla---ValAsnAspLeuIleSerPro 484
Db 3132 ACGGGTCCCGTCCCGCGGTCGGACCGCGGTTCATCCGCACTCGCGCTCGCC 3191
Qy 485 ProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyPheThrHisAlaAsp 504
Db 3192 GCGCGCGACGCCACC-----GACTCGCGCCACGGTCGCAACAGCTCGAC 3233
Qy 505 ValThrAlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSerVal 524
Db 3234 GTCACTCTCGGTCCCGCGGATCCGCGCGGCGAGGGCC---ACCGCGCAGACCTGGGTC 3290
Qy 525 AspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSerSer 544
Db 3291 GAT-----GAACCGCGCGCC 3305
Qy 545 GlnProAsnAspArgHisValAlaPheSerThrAspGlyLysAsnTrpPheGlnGly 564
Db 3306 GACGGCGCGCGCGCTTCCACGCTCCACCGCGTCCGCGACGCCCGCTGCACGCTGCAC 3365
Qy 565 SerGluProGlyGlyValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArg 584
Db 3366 GCGGAG-----GGGGTCTCCGCGCGCGCGGTG---CCCCAGCCGGAAGCGTGCAC 3416
Qy 585 PheValTrpAlaProGlyAspProGlnProValValValValValValPheGlyAsn 604
Db 3417 ACGGCTGGCCCG-----CGGGC-----GCGGTGCGCGGACGGCTCCCGGG 3464
Qy 605 SerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsn 624
Db 3465 GCGTGGCGACGCGCGGACCAAGTCTCTCGTCCGAGCCGAGTGCAGAC----- 3512
Qy 625 ProLysThrPheTyrAla-----LeuSerAsnGlyThrPheTyrArgSerThrAsp 641
Db 3513 CTTGACGGCTTCGTGGCACACCCCGACCTCTCGACGCGGTCTTCTCCGCGGTGCGGCGAC 3572
Qy 642 GlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyVal 661
Db 3573 GGG-----AGCGCGACGCGACCGGATG---GCGGACCTCGCGTGCACGCTCGGACG 3625
Qy 662 MetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeu 681
Db 3626 CACCGTGTGCGCGCTCGCTCACCCTCGGACAGTGTGTGTCGAGCTCGCGGCTT 3685
Qy 682 TyrHisSerThrAnGlyGlySerTrpSerAlaIleThrGlyValSerAla--- 700
Db 3686 CGACGGTTCGCGGAATGCGGT-----GCTCACCCTCGGAGTCCGTGACGCTGGG 3733
Qy 701 ---ValAsnValGlyPheGlyLysSerAlaProGlySerSerTrpProAlaValPheVal 719
Db 3734 CGAGGTGCGTGCAGCGGATCCGACGAGTCCGCTGCTCGCTCGGTTCGAGTGT 3793
Qy 720 ValGlyThrIleGlyGlyVal-ThrGlyAlaTyrArgSerAspAspCysGlyThrThrTr 739
Db 3794 GCGGGTGGCGGAGCCCACTACGAGGTGCC-----GACGAGTGGCCCGGAGGCTA 3844
Qy 739 pValLeuIleAsn-----AspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAl 755
Db 3845 CACCCCTCATCCCGCCACACACCCGACGACCCCGACGACCCCAACCAACCCCAACAC 3904
Qy 755 alleThrGlyAspHisAlaAsnLeuArgArgVal-----TyrIle 768
Db 3905 ACCCACGACGACCCACACACAAACACACGCGTCTCACCGCCCTCCAAACACCACTCAT 3964
Qy 768 eGlyThrAsnGlyArgGlyIleValTyrGlyAspIle-----GlyGlyAlaProSe 785

Db 3965 CACCAACCAACACCCCTCATGTCACCAACCAACCAACCAACCAACCAACCAACCAAC 4024
Qy 785 rGly-SerProSerPro-----SerValSerProSerAlaSerProSerLeuSerPro- 802
Db 4025 CGGCTCACCGCGACCGCAAAACGAACACACCCCGCGCATCCACCTCATCGAACCAC 4084
Qy 803 -----SerProSerProSerProSerProSerProSerProSerProSerPro- 814
Db 4085 CCACCCCAACACCCGACCTCCGCTCAGCACTCAGCACTCAGCACTCAGCACTCAGCACT 4144
Qy 815 -----SerProSerProSerProSerProSerProSerProSerProSerPro- 831
Db 4145 CCTCACAACCAACACCCCTCCACACCCCGCACCTCAGCACTCAGCACTCAGCACTCAG 4204
Qy 831 erProSerProSerArgSerProSerProSerProSerProSerProSerProSerPro- 849
Db 4205 CACCACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 4264
Qy 850 -----ProSerProSerProSerProSerProSerProSerProSerProSer- 866
Db 4265 CGGCTCGGCAACCTCGCGGCTCTCGCGCGCACCTCAGCACTCAGCACTCAGCACTTAC 4324
Qy 866 erSerPro 868
Db 4325 CTTCTCC 4332

RESULT 9

US-09-252-991A-13773/c
; Sequence 13773, Application US/09252991A
; Patent No. 6551795

GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 13773
; LENGTH: 4131
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13773

Alignment Scores:

Pred. No.:	8,98e-06	Length:	4131
Score:	288.50	Matches:	246
Percent Similarity:	31.64%	Conservative:	109
Best Local Similarity:	21.93%	Mismatches:	376
Query Match:	5.62%	Indels:	392
DB:	4	Gaps:	56

US-09-917-376-1 (1-957) x US-09-252-991A-13773 (1-4131)

Qy 37 IleAlaIleThrAlaSerProAlaHisAlaAlaThrThrGlnProTyrThrTrpSer--- 55
Db 4113 GTCAACGTGACCGCGACGCGCGCGCAATACCGCGCTCCGCTACCAACCGGTG 4054
Qy 56 -----AsnValAlaIleGlyGlyGlyPheValAspGly 67
Db 4053 GATTCTCTGCTCGCGTCCGATCCCGAGGTGGATCCGAGCAACCGGTTCGATGATCAGCGG 3994
Qy 68 IleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyMet 87
Db 3993 -----ACGCGGACCGCGGCAACACC 3973
Qy 88 TyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsn 107

3972 ATCATCATCACCAGTGGCAACGCAACCCGATGGCCAGGTCTACCGCCGACGCGCGT 3913
Db
108 AsnTrpGlyTyr-----AsnGlyValVal-----SerIleAla 118
Qy
3912 AACTGGTCTTCACTCCAGGCATCCCGTCCCGATGGCATGGTCAACGCTGGTGGCG 3853
Db
119 AlaAspProIleAsnThrAsnLysValTrpAla-----AlaValGlyMet 133
Qy
3852 CGCAGCCCAAGCAATGTCGACAGTGGCCGCGCGGTGATCACTGGATGGCGTGGCCCCG 3793
Db
134 TyrThrAsnSerTrpAspProAsnAspGlyAla---IleLeuArgSerSerAspGly 152
Qy
3792 CGCGCGCGGTGATCGATCCGACACAGCGCACCGAGATAGCGGTACCGGAGGCGCGC 3733
Db
153 AlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGlyArgGly 172
Qy
3732 GCGAGGTGATCCTCACCGAT-----GGCGCGGCAACCCG----- 3697
Db
173 MetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAlaPro 192
Qy
3696 ATCGCCGAGGCCACCGCC----- 3679
Db
193 SerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThrAsn 212
Qy
3678 GACGGCAGCGCAACTGGACGTTCAACCCCG----- 3649
Db
213 PheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGlnSerAsp 232
Qy
3648 -----GGCACCCGCTGGCCACAGCGCACCGTGATCAACGCC----- 3613
Db
233 IleGlnGlyValValTrpValAlaPheAspLysSerSerSerLeuGlyGlnAlaSer 252
Qy
3612 -----GTGGCCAGGACCCCGCGCAATACCAAGCGGTCCGCCAGC 3571
Db
253 LysThrIlePhe-----ValGlyValAlaAspProAsnAsnProValPhe 267
Qy
3570 GTCACCGTCGATCCATCCATCCCGCGCGCGCGGTGATCAATCCGAGCAACGAGTGTCTC 3511
Db
268 TrpSer----- 269
Qy
3510 ATCAGCGGTACGGGGAAGCGGGGCCACCGTGATCTCAACGAGGCAACGCGCAACCCG 3451
Db
270 -----ArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaPro--- 283
Qy
3450 ATCGCCGAGTACCGCGGACGCGCGCAAGTGGCTTTACGCCCGCCACGCGGTG 3391
Db
283 ----- 283
Qy
3390 GCCAATGGCACCGGTGATCAATGGCTGGCCGACGAGCGCGCGCAACACAGCAGTCCC 3331
Db
284 -----ThrGlyPheIleProHisLysGlyValPheAspProValAsnHis 298
Qy
3330 ACCAGCGCCACCGTCGACTCGTGGCGCCAGCAGCCCGGTGATCGATCCGAGCAACGGT 3271
Db
299 ValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyr-----AspGlySerSer 315
Qy
3270 AGCGTATCGCCGCTACCGCGAGGTGGTGCCACCGTGATCTCAACGAGCGCAACGGC 3211
Db
316 GlyAspValTrpLysPheSerValThr---SerGlyThrTrpThrArgIleSerProVal 334
Qy
3210 AACCCGATCGCGCGGTCACCGCGATGGCAGCGGCAACTGGAGCTTCACGCGCGGCAGC 3151
Db
335 Pro-----SerThrAspThrAlaAsnAspTyrPhe 344
Qy
3150 CCGCTGTCCAATGGCACGCGTGGTCAATGGCGGTGCCCGAGCAGCTGCGCGCAACACGAGC 3091
Db
345 GlyTyrSerGlyLeuThrIleAspArgGlnHisPro----- 356
Qy
3090 GGCCCGCCAGCACACCGTGGTACTCGTGGCGCCCGCGCGGTGATCGACCGGAGC 3031
Db
357 AsnThrIleMetValAlaThrGlnIleSerTrpTrpProAspThrIleIlePheArgSer 376
Qy
3030 AACCGCAGCGTGTGATCGCGGTATCCGCGGAAGCGGTGCGAGTGTGATCTCTC----- 2980
Db

377 ThrAspGlyGly-----AlaThrTrpThrArgIleTrpAspTrpThr 390
Qy
2979 ACCGATGGCGCGCAACCCGATCGGCAGGCAACCGCGGATGGCAGCGCACTGGAGC 2920
Db
391 SerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAla----- 405
Qy
2919 TTCACCCCGGCGACGCGCTGGCCACCGCACGGTGTCAATCGCGTGGCCAGGATCCG 2860
Db
406 -----GluProTrpLeuThrPhe-----GlyValGlnProAsnProPro 418
Qy
2859 GCCGGCAATACACGCGCGCCGACACGACACCGGTGGACGCGGTGCCCGCCGCCCGC 2800
Db
419 ValProSerProLysLeuGlyTyrMetAspGluAlaMetAlaIleAspProPheAsnSer 438
Qy
2799 GTGCTCAACCCGAGCAACCGC----- 2779
Db
439 AspArgMetLeuTyrGlyThr-----GlyAlaThrLeuTyrAlaThrAsnAspLeu 455
Qy
2778 ---AGCGTGTATCGCGGTACCGCGGAAGCCGCGCACCGGTGATCCTCAC----- 2731
Db
456 ThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGlu 475
Qy
2730 -----GACGCGCGCGCAAC-----CCGATCGCGCAGGTCAACGCCGAC 2692
Db
476 ThrAlaValAsnAspLeuIleSerProSerGlyAlaProLeu----- 490
Qy
2691 GGCAGCGCAACTGAGCTTCACGCC-----GGCACGCGCTGGCCAAACGGTCCGTTG 2638
Db
491 IleSerAlaLeuGly---AspLeuGlyGlyPheThrHisAlaAspVal---ThrAlaVal 508
Qy
2637 ATCAATGCGTGGCGCGAGGACGCGCGCAACACGCGCGCGCGCCAGCACCGCGTG 2578
Db
509 ProSerThrIlePheThrSerProValPhe-----ThrThrGlyThrSerValAsp--- 525
Qy
2577 GACTCGGTAGCCCGCCACCCCGGTGTCTGATCGAGCAACGGTACGGTGTACCGGT 2518
Db
526 TyrAlaGluLeuAsnProSerIleIleValAlaGlySerPheAspProSerSerGln 545
Qy
2517 ACCGCGCAAGCGCGCGCCACCGGTGATCTCACGAGCGCGCGCGCAACCGCATACG 2458
Db
546 ProAsnAspArgHisValAlaPheSerThrAspGlyLysAsnTrpPheGlnGlySer 565
Qy
2457 GCC-----ACGCGCATGGCAGCGCAACTGG-----AGCTTC 2425
Db
566 GluProGlyGlyValThrThrGlyThrVal-----AlaAlaSerAlaAspGlySer 583
Qy
2424 ACTCGGCGCACCGCTGACCAACCGCACGGTGTCAATGCGGTGGCCGACGACGCGCC 2365
Db
584 ArgPheValTrpAlaPro----- 589
Qy
2364 GGCAACACCGCGTCCGTCAGCACACAGTGGACGCGGTGGCCCGCCACCCCGGTG 2305
Db
590 -----GlyAspProGlyGlnProValVal 597
Qy
2304 ATCGACCGGACCAATGGTGTCAAACTCAGCGCACCGCGCAACCGCGGTCCGGGTGATC 2245
Db
598 TyrAlaValGlyPheGlyAsn-----SerTrp 606
Qy
2244 CTCACCGATGGCAATGGCAACCCGATCGCGCAGACCCCTCGCGCAGCGGTAGCGCAACTGG 2185
Db
607 AlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLys 626
Qy
2184 ACCTTCACCGCGCACGCC----- 2164
Db
627 ThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPhe 646
Qy
2163 -----CTGGCCACCGCACCGGTGTCAACGCC-----GTGGCCAG 2128
Db
647 GlnProValAla---AlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAla 665
Qy
2127 GACCGCGCGGCAATACACGCGGTCCGCGCACGACCGGTGGATACGCGTGGCGCGCGCC 2068
Db

```
Qy 666 valProGlyysGluGlyAspLeuTrpLeuAlaSerSerGlyLeuTrpHisSerThr 685
Db 2067 ACACCG-----GTGATCAATGCCAGC 2047
Qy 686 AsnGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGlyPhe 705
Db 2046 AACCGC-----AGCGTGATCACCGCACCGCC-----GAGTTCGGC--- 2011
Qy 706 GlySerAlaProGlySerSerTrpProAlaValPheValValGlyThrIleGlyGly 725
Db 2010 -----GCCAAGTGATCTCTCACCGACGCAACGCAAC 1978
Qy 726 ValThrGlyAlaTyArgSerAspCysGlyThrThrTrpValLeuIleAsnAspAsp 745
Db 1977 CCGATCGCGGACACCAACCGCGGACGCGAGTGGC---AACTGGACCTTCACCCCGGCGACG 1921
Qy 746 GlnHisGlnTyClyAsnTrpGlyGlnAlaIleThrClyAspHisAlaAsnLeuArgArg 765
Db 1920 CCGCTGGCCAACGGTACGGTGATCAACCGCGTCCGGAAGACCGCGC----- 1873
Qy 766 ValTyrlleGlyThrAsnGlyArgGlyIleValTyArgGlyAspIleGlyGlyAlaProSer 785
Db 1872 -----GGCAACGCCAGCGGT---CCGGCC 1852
Qy 786 GlySerProSerProSerValSerProSerAlaSerProSerLeuSerProSerProse 805
Db 1851 AGCACCAACCGGTGACTCGGTGGCGCGTCCGCT---CCGCTGTGTGATCATCAGCGCGCAC 1795
Qy 805 rProSerSerProSerPro-----SerProSerProse 817
Db 1794 GCGCGCTGTGATCACCGCGCACCTGATCGCGGGGTTCGCTGGACGCGCGCGCAAC 1735
Qy 817 rSerSerProSerSerSerProSerProSerProSerProSerProSerArgSe 837
Db 1734 GACACCGCAACCCGATCACGTCACGTCGACGCGCGCGGCACTTCACCTGCCGCTTC 1675
Qy 837 r-----ProSerPr 840
Db 1674 GCGCGCGCTGATCACCGCGCAACTGATCGCGGGGTTCGCTGGACGCGCGCGCAAC 1615
Qy 840 oSerAlaSerProSerSerSerProSerProSerProSerSerSerProSerSer---Se 859
Db 1614 GTCAGCGCGCGCGCCACCATCAACGCGCGCGGACCTGGCGCGCGCGGACCATCAGCGTCCG 1555
Qy 859 rProSerProThrProSerSerSerProValSerGlyValValSerValGlnTyrlsAs 879
Db 1554 GAACCGCGCGATACCTGGATCAACGCGCGGAGATCGGGACGCGC-ATCCAGGTC----- 1501
Qy 879 nAsnAspSerAlaProGlyAspAsnGlnIleLysProGlyLeuGlnValValAsnThrGl 899
Db 1500 -----GATGTGACGGTCCGTCGACCATGACGTCGCGC----- 1468
Qy 899 ySerSerSerValAspLeuSerThrValThrValArgTyTrpPheThrArgAspGlyGl 919
Db 1467 -----CAGGTGTGTACGCTCAAGTTC---GCCGGGCGACGCGCTA 1430
Qy 919 ySerSerThrLeuValTyAsnCyAspTrpAlaAlaIleGlyCysGlyAsnIleArgAl 939
Db 1429 CGAGGCGGAGGTCCCATACCTCACCAGCGCGGACATCGCGCGCGGCAACCTGACCTT 1370
Qy 939 aSerPhe-----GlySerValAsnPro-----AlaThrProThrAl 951
Db 1369 GACCTGACGCTCCCGCGGCGATGGCCCGTTCGCGGAGGCGCGCTCGACCGCTCACCGC 1310
Qy 951 aAsp 952
Db 1309 CGAC 1306
```

RESULT 10

```
US-09-410-551B-18
; Sequence 18, Application US/09410551B
; Patent No. 6503737
; GENERAL INFORMATION:
```

```
; APPLICANT: KOSAN BIOSCIENCES, INC.
; APPLICANT: REEVES, CHRISTOPHER
; APPLICANT: CHU, DANIEL
; APPLICANT: KHOSLA, CHAITAN
; APPLICANT: SANTI, DANIEL
; APPLICANT: WU, KAI
; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA
```

```
; FILE REFERENCE: 30062-20026.00
```

```
; CURRENT APPLICATION NUMBER: US/09/410,551B
```

```
; CURRENT FILING DATE: 1999-10-01
```

```
; PRIOR APPLICATION NUMBER: US 60/139,650
```

```
; PRIOR FILING DATE: 1999-06-17
```

```
; PRIOR APPLICATION NUMBER: US 60/123,810
```

```
; PRIOR FILING DATE: 1999-03-11
```

```
; PRIOR APPLICATION NUMBER: US 60/102,748
```

```
; NUMBER OF SEQ ID NOS: 72
```

```
; SOFTWARE: FastSeq for Windows Version 4.0
```

```
; SEQ ID NO 18
```

```
; LENGTH: 4571
```

```
; TYPE: DNA
```

```
; ORGANISM: Artificial Sequence
```

```
; FEATURE:
```

```
; OTHER INFORMATION: Description of Artificial Sequence: DNA encoding synthetic
```

```
; OTHER INFORMATION: PKS synthase fragment
```

```
; FEATURE:
```

```
; NAME/KEY: CDS
```

```
; LOCATION: (9)..(4559)
```

```
; US-09-410-551B-18
```

```
Alignment Scores:
```

```
Pred. No.: 1,01e-05 Length: 4571
```

```
Score: 288.50 Matches: 227
```

```
Percent Similarity: 34.22% Conservative: 121
```

```
Best Local Similarity: 22.32% Mismatches: 390
```

```
Query Match: 5.62% Indels: 283
```

```
DB: 4 Gaps: 45
```

```
US-09-917-376-1 (1-957) x US-09-410-551B-18 (1-4571)
```

```
Qy 9 LeuThrMetArgSerArgArgLeuValSerLeuLeuAlaAlaThrAlaSerPheAlaVal 28
Db 1662 CTGACCGAACACGACGACCGCTGCGCGCTACCTGCGCGCGGCGGATATA 1721
Qy 29 AlaAlaAlaLeuGlyValLeuProIleAlaIleThrAlaSerProAlaHisAlaAlaThr 48
Db 1722 CGGCTGTGGCATCGACGCTG-----GCGGTGACACGCTCGGT-GTTCGAGCAGCGCGC 1774
Qy 49 ThrGlnProTyr----- 52
Db 1775 CGTACTCTTGGAGATGACACCGTCAACCGGACCGCGGTGACCGACCGGATCGTGT 1834
Qy 53 -----ThrTrpSerAsnValAlaIleGlyGlyGly-----PheVal 65
Db 1835 TGTCTTCCCGCGAGGCGGTGCGAGTGGCGGATGGCGAGTGCAGTGCAGTGCAGTTCGTC 1894
Qy 66 AspGlyIleValPheAsnGluGly-----AlaProGlyIleLeuTyValArgThrAsp 83
Db 1895 GGTGCTGTTCGCGAGCGGATGGCGGCGGTGTCGCGCGGTTCGCGAGTTCGT---GGA 1951
Qy 84 IleGlyGlyMetTyArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrp 103
Db 1952 CTGGGATCTGTTCAGGTTCTGGATGATCCGCGCGTGG-TGGACCGGTTGATG---TGG 2007
Qy 104 ValGlyTrpAsnAsnTrpGlyTyAsnGlyValValSerIleAlaAlaAspProIleAsn 123
Db 2008 TCCAGCCC-GCTTCTCTGGCGATG-----ATGTTTCCCTGGCGCG----- 2048
Qy 124 ThrAsnLysValTrpAlaAlaValGlyMetTyThrAsnSer---TrpAspProAsnAsp 142
Db 2049 -----GTGTGGCAGCGCGCGGTGTGCGCGCGGATGCGGTGATCGCGCATTCGCAG 2099
```


Qy	143	GlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGln-----	156
Db	2100	GGTGAGATCGCCGCGAGCTTGTGTGGGGGTGCGGTGTCTACTACGCGATGCGCGCCGATC	2159
Qy	157	IleThrProLeuProPheLeysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArg	176
Db	2160	GTGACCTTGGCGACCCAGCGATCGCCCGGGGCTGTGGCGGCGCGGGCGGATGGCATCC	2219
Qy	177	LeuAlaValAspProProAsnAspAsnIleLeuTyPheGlyAlaProSerGlyLysGly	196
Db	2220	GTGCGCCCTG---CCCGCGCAGATGTCGAGCTGGTCAGCGGGGCC-----	2261
Qy	197	LeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspVal	216
Db	2262	---TGGATCGCGCCCAACAGCGGCGCGCTCCACCGTGATCGCGGCGACCCCGGAACGC	2318
Qy	217	GlyThrTyrlleAlaAsnProThrAspThrGly-----	228
Db	2319	GTGACCATGTCTCCACCGCTCATAGAGCAAGGGGTGCGGTGCGGGGATCACCGTC	2378
Qy	229	---TyrGlnSerAspIleGlnGlyValValTrpVal-----	244
Db	2379	GACTATGCTCGCACACCCCGCAGCTCGAGCTGATCGCGACGAACCTACTCGACATCACT	2438
Qy	245	SerSerSerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsn	264
Db	2439	AGCGACAGCAGCTCGCACAGCCCGCTCGTCCGCTGGTGTGCGACCGTGGACGGCACC---	2495
Qy	265	ProValPheTrp-----SerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAla	282
Db	2496	-----TGGGTGCGACAGCCCGCTGGACGGGGAGTACTGTGACCGGAACCTTGCCTGAA	2546
Qy	283	ProThrGlyPheIleProHisLysGlyValPheAspProValAsnHisValLeuTyrlle	302
Db	2547	CCGGTGGTTTCACCCCGCGCTGACCGCATGTCAGCGCCAGCGGCGACACCGTGTTCGT	2606
Qy	303	AlaThrSerAsnThrGlyGlyProTyTrpAspGlySerSerGlyAspValTrpLysPheSer	322
Db	2607	GAGGTACGCCAGCCCGGTGTTGTCAGCGCATGGACGACGATGCTGTGTCAGGTGTC	2666
Qy	323	Val-----ThrserGlyThrTrpThrArgIleSerProValProSerThrAspThr	339
Db	2667	ACGCTGCTGCTGACGACGCGCGACGCCACCGCATG-----CTCACCGCCCTG	2714
Qy	340	AlaAsnAspTyrrPheGlyTyrrSerGlyLeuThrIleAspArgGlnHisProAsnThrIle	359
Db	2715	GCACAGCCTAT-----GTCCAGCGCGTCACCGTCGAC-----	2747
Qy	360	MetValAlaThrGlnIleSerTrpTrpProAspThrIleIlePheArgSerThrAspGly	379
Db	2748	-----TGGCGCGCATCTCTC-----	2762
Qy	380	GlyAlaThrTrpThrArgIleTrpAspTrpThrSerTyrrProAsnArgSerLeuArgTyrr	399
Db	2763	GGCACCACCAACACCGGGTACTGGACCTTCCGACCTACGCTTCCAAACACCAACCGCGTAC	2822
Qy	400	ValLeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnProVal	419
Db	2822	-----	2822
Qy	420	ProSerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAsp	439
Db	2823	-----TGGCTCGAGTCGGCACGC-----	2855
Qy	440	-----ArgMetLeuTyrrGlyThrGlyAlaThrLeuTyrrAlaThrAsnAspLeuThrLys	457
Db	2856	CGCGGGCCACCCGTGTGGGCTCCGGTATCGCCCTCGCCGGG-----	2897
Qy	458	TrpAspSerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluThrAla	477
Db	2898	-----TCGCGCGCGGGGTTCACGGGT-----	2921
Qy	478	ValAsnAspLeuIleSerProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeu	497

[illegible]

ORGANISM: Mycobacterium tuberculosis
OTHER INFORMATION: H37Rv
US-09-103-840A-1

Alignment Scores:

Pred. No.: 0.0493 Length: 4411529
Score: 282.50 Matches: 278
Percent Similarity: 32.66% Conservative: 111
Best Local Similarity: 23.34% Mismatches: 377
Query Match: 5.50% Indels: 432
DB: 3 Gaps: 59

US-09-917-376-1 (1-957) x US-09-103-840A-1 (1-4411529)

Qy 4 SerGluAsnIleArgLeuThrMetArgSerArgArgLeuValSerLeuLeuAlaAlaThr 23
Db 372877 TCCAGCGACTCCAGGCTTCGGCAACTCGACCGGGGACTGTC----- 372836
Qy 24 AlaSerPheAlaValAlaAlaLeuGlyValLeuProIleAlaIleThrAlaSerPro 43
Db 372835 -----GTCCGGCTTCTTCA-ATAGCGCGCTGGCAGCGCCT 372801
Qy 44 AlaHisAlaAlaThrThr--GlnProTyr-----ThrTyrSerAsnValAla 58
Db 372800 CGGGCTTCGGCAACTCGGGGGCGCCGCTGGCTTTATCAACCTGTCTCC-ACAAAG 372742
Qy 59 IleGlyGlyGlyPheValAspGlyIleValPheAsnGlyAla-----ProGly 76
Db 372741 TCGGGAATGTCGGCTTCTTC-----AACGTGGCGCGCTGGGATCGGGT 372697
Qy 77 IleLeuTyrValArgThrAspIleGlyMetTyrArg-TripAspAlaAlaAsnGlyAr 96
Db 372696 GTGGCGAATGTGGCAACACCACTCGGTATCTACAACTGGCGCAGCTCGGACTCTCG 372637
Qy 96 gTrrIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyrAsnGlyValValse 116
Db 372636 ACGCCGCCCTTAACCTCG-----GGTTGGCAA----- 372609
Qy 116 rIleAlaAlaAspProIle-----AsnThrAsnLysVal----- 127
Db 372608 -----ATATCGGAACCAATATTCGGCGCTGTGCGCGACGGCGGTACTCGGGCTATT 372553
Qy 128 -----TrpAlaAlaValGlyMetTyrTrpAsnSerTrpAspProAsnAspGlyAlaIleLe 146
Db 372552 AACTGGGCTTGCCCAACACCGCAACTCAACGTGGGCTTCGCAAGTCTCGGGGCTTT 372493
Qy 146 uArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyG 166
Db 372492 AACTTCG----- 372486
Qy 166 yAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnI 186
Db 372485 -----GCGCGCCACCACTCGGCCCAACCAAC 372460
Qy 186 eLeuTyrPheGlyAlaPro--SerGlyLysGlyLeuTrpArgSer----- 200
Db 372459 GTGGCATCGGGAACACCGGAATCTTCGATGTGGCTTCGGCACTCGGCGCAGCTACAAC 372400
Qy 201 -----ThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspVa 216
Db 372399 ATCGGCTTCGGAACCTTCGCGAGCAGCAACCTGGCTTCGGCAACTTCG-----GC 372349
Qy 216 lGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyVa 236
Db 372348 AGCTACAAACATCGGCTTCGGCAACGTTCGGCAACG-----AC 372313
Qy 236 lValTrpValAlaPheAspLysSerSer----- 245
Db 372312 AATCTGGTTCCTTAACCGGGCGGGCGCAACATCGGCTTTTGGAAACACCGGAGCAAC 372253
Qy 246 -----SerSerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAs 264
Db 372252 AATGTCGGCTTCGGGAACACCGGCGCAACATATGTCGGCATCGGGCTCAGCGGCACCGGA 372193

Qy 264 nProValPheTrpSerArgAspGlyGlyAla----- 274
Db 372192 C-----AGATCGGTTTCGGCAGCTTCAACTCGGGCAGCGGAACATCGGC 372148
Qy 275 -----ThrTrpGlnAlaValProGlyAlaProThrGlyPh 286
Db 372147 CTGTTCAACTCGGCGAGCAACAACTCGGATTCCTCAATCCGGCAGCGCACTTCGGC 372088
Qy 286 eIleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAs 306
Db 372087 -ATCGCAAACTCGGCGAGCTTCAAC-----ACTGGCATCGGAAA 372050
Qy 306 nThrGlyGlyProTyrAspGly-----SerSerGlyAspValTrpLysPheSerValTh 324
Db 372049 CACCGCAACCAATAATCCGGCTTATCAACTCCGGCGAC-----GTCAA 372005
Qy 324 rSerGlyThrTrp-ThrArgIle-----SerProValProSerThrAspThrAlaAsnA 342
Db 372004 CACGGCGCCTTCAACCCGGCGAGCTTCAACACCGGTAGCTTCAACACCGCAGCTTCAA 371945
Qy 342 sPtyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValA 362
Db 371944 CACCGTGGCT---TCAATCCGGGCAATACCAACACCGGCTACCTCAACTTGGCAACTA 371888
Qy 362 laThrGlnIleSerTrpTrpProAspThrIleIle-----PheArgSerThrAspGlyG 380
Db 371887 CAACACCGGATCGCCAAACACCGCGGAGCTTGACACCGGGGCTTTCATCACCGGAAACTA 371828
Qy 380 lyAlaThrTrp-----ThrArgIleTrpAspTrpThrSerTyrProA 394
Db 371827 CAGCAACGGGTTGTTCTTAAGCGCGGATTAACAGGCGCTGG----- 371787
Qy 394 snArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrpLeuThrPheGlyValG 414
Db 371786 -----TCGGCTCAACCTGGTGCATCATATGCTCT 371756
Qy 414 lnPro-----AsnProProValProSerP 422
Db 371755 CCCATAAGCTCGCGGTGAATATTCCTCATCGATATCCGATACCGCTCGGCCGCA 371696
Qy 422 rLysLeuGlyTrp----- 426
Db 371695 CATCACCTTATGGGCGTCAAGATTCCGCCACCGCGATATCGTCTTTCGTCAATAGC 371636
Qy 427 -----MetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyr--G 444
Db 371635 GGGCCAGCGAGCCCACTTTGGCCCCATTAACATTCGGAACATCACGGTTGTGGCCCCAC 371576
Qy 444 yThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlnI 464
Db 371575 GACGAGTCCCATAGAGGGCGGATACCGCGATACCATTAACGTGGCGGTGGGCCCAT 371516
Qy 464 eHisIle-----AlaProMetValLysGlyLeuGluThrAlaValAsn-- 479
Db 371515 TAGGATCCGCTCATCAGTATCCCGCGCGCCAGGTTTCGGAACCTCGACCAACCC 371456
Qy 479 ----- 479
Db 371455 GTCGTGAGTTTCTTAATACCGCGCGCGCTCGGGCTTCGGCAACTTCGGCGG 371396
Qy 480 -----AspLeuIleSerProProSerGlyAlaProLeuIleSerAl 493
Db 371395 CGCAATTCGGGCTTTTGGAACTTGGCTTCGGCGACCTCGGGGCGTTCCTCAA 371336
Qy 493 aLeuGlyAspLeuGly-----GlyPheThrHi 502
Db 371335 CGTCGGCGCTTGGGATCAGTCTGGGCAACGTGGGCAACCGCTCTCGGGGTTCACAA 371276
Qy 502 sAlaAspValThrAlaValProSerThrIlePheThrSerProValPheThrGlyTh 522
Db 371275 CACCAGCAGCTCGGACCTCGGCGAGCGCGGCTTCAATTCAGGCTTGGCCCAACATCAGCAC 371216

QY 522 rSerVal----- 524
Db 371215 CAGTATCGCGGCTGTGCGCAGACGACCGGCATGCTCCTCAACCTGGGCTTGGC 371156
QY 525 -AspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSerSe 544
Db 371155 AAACACCGGACCTCAACGTCGGCATTT-----GC 371126
QY 544 rGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrp----- 561
Db 371125 AAACCTCGGCGACTACACATCGCTTTCGAACCTCGGCGAGCCCACTTCGGCAGCGC 371066
QY 562 ----PheGlnGlySerGluProGlyGlyValThr----- 572
Db 371065 CAATATCGGTGGCAACACATCGCGCGCAACACCGCAATATTCGACATCGTTTGGC 371006
QY 573 -----GlyGly-ThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaProG 590
Db 371005 AAATCGGGCAGTTACACATCGCTTCGGAACCTTCGGCGATGACAACTGGGCTTC-G 370947
QY 590 LysAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerG 610
Db 370946 GAAACCTCGGCAGC-----TACACGTCGGCTTCGGAACCTTCGGCAACGACCAACC 370896
QY 610 LnglyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyrA 630
Db 370895 TGGGCTTC--GCCAACACCGGC----- 370876
QY 630 lalLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProValA 650
Db 370875 ----AGCAACAATATCGGCTTCGGCAACACCGGCAGC-----AACAAATATCG 370833
QY 650 laAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaValProGlyLysG 670
Db 370832 GCATTCGGCTTCAGGGCGAGCGCAGATCGGG-----TTCGGCTCCTCAATTCGCGCA 370779
QY 670 luGlyAspLeuTrpLeu-----AlaAlaSerSerGlyLeuTyrHisSerThrA 686
Db 370778 GCGGAAACATCGCTTGTTCACCTCGGCGAGCGGAAACATCGCTTTTCAACTCGGCGCA 370719
QY 686 snGly-----GlySerSerTrpS 692
Db 370718 ACGGAAACGTTGGCATCGGCAACACCGGCACCGCAAACTTCGGGCTTGGAAACACCGCGCA 370659
QY 692 erAlaIleThrGlyVal-----SerSerAlaValAsnValGlyPheGlyLysSerAla- 709
Db 370658 GCACCAACACCGCTTCTCACTTCGGCGAGCTCAATACCGTATCGGCAACACCGCGCA 370599
QY 710 -----ProGlySerSerTyrProAlaValPheValValGlyT 722
Db 370598 GCCTTCAACACCGCGAGCTTCAATCGGCGGATTCACACACCGGGGATTTCAACCCAGCGCA 370539
QY 722 hrIleGlyGlyValThrGlyAlaTyrArgSerAspAspCysGlyThrTrpValLeuI 742
Db 370538 GCTACAAAC-----ACGGGACTCGGAACACCGCGGATGTTCACAC----- 370498
QY 742 leAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaA 762
Db 370497 -----GGCGCTTCATCTCGGCGAGCTACAGCA 370470
QY 762 snLeuArgArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGlyG 782
Db 370469 AC-----GGGTTCTTGTGGAGTGAATATTCAGGGGCTCATTCGCT 370428
QY 782 lyaAlaProSerGlySerProSer-----ProSerValSerProSerAlaSerProS 799
Db 370427 T-GCAGCGCGCGCTAGCGATTTCGGAATTCGCGCAATCGCCCTAACCTTTG-GCGTCGACATCCCGAT 370370
QY 799 erLeuSerProSerProSer-----ProSerSerPro----- 810
Db 370369 ACATATACCCATCAACATCGACCGGGGTCGTACCCCTCCAGCGGCTTCAGCATCGTAGC 370310
QY 811 -----SerProSerProSerProSerSerSerProSerSerSerProSerP 826

Db 370309 TGCAGAAATAATATCGACTTCAACCCCATCATCCCGACCATATATCACCTTGGC 370250
QY 826 ro-----SerProSerProSerProSerProSerProSerProSerProSerProSerAla 843
Db 370249 CACGGCGCGCATCACCGTGGCGGACCCA-----CCACCTCGATCGGTATCACCG----- 370200
QY 843 erProSerProSerProSerProSerProSerProSerProSerProSerProSerP 860
Db 370199 --CCAGCGCGGTATCGCTCCATCACCATCCGATCATCGACATTCGCGGACATCGGG 370142
QY 860 roSerProThrProSerSerSerProValSerGlyGlyValLysValGlnTyrLysAsnA 880
Db 370141 CTTGGCAACTCG-ACCACTAGTCGTCGCGGG-----TTCTTCAACT 370098
QY 880 sn-----AspSerAlaProGlyAspAsnGlnIleLysProGlyLeuGln----- 894
Db 370097 CCGAGCGGCGCAGCGCTCGGGCTTTTGAACGTGGTTCGCGCGCTCAGGGATTCGG 370038
QY 895 -----ValValAsnThrGly----- 899
Db 370037 GTTATCTCAATGTCGTCGCTGGGTGGGTGTGACTAAACGTGGGTACACCGCTCTCGG 369978
QY 900 -----SerSerSerValAspLeuSerThrValThrValArgTyrTrpPheThrA 916
Db 369977 GTTTCTACAACGCGAGCGCTTGGACCTCGTACCGCGCGCTTTCCTCCGCTCTCATCG 369918
QY 916 rGAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAlaIleGlyCysGlyA 936
Db 369917 GCGHAC-----GGTATGGGACGATGACTCTGAACCTTGGCTGGCAACCTGGGACGACATA 369861
QY 936 snLeuArgAlaSerPheGlySer 943
Db 369860 AC-----GCCGGCTTCGGCAAC 369844
RESULT 14
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2
Alignment Scores:
Pred No.: 0.0552 Length: 4403765
Score: 281.50 Matches: 278
Percent Similarity: 32.86% Conservative: 111
Best Local Similarity: 23.34% Mismatches: 377
Query Match: 5.48% Indels: 432
DB: 3 Gaps: 59
US-09-917-376-1 (1-957) x US-09-103-840A-2 (1-4403765)
QY 4 SerGluAsnIleArgLeuThrMetArgSerArgLeuValSerLeuAlaIleThr 23

Db 372934 TCACGCGACTCCAGGCTTCGGCAACTCGACCGCGGACTGCTC----- 372893
Qy 24 AlaSerPheAlaValAlaAlaLeuGlyValLeuProleAlaIleThrAlaSerPro 43
Db 372892 -----GTCGGCTTCTTCA-ATAGCGGCGCTGGCAGCGCT 372858
Qy 44 AlaHiAlaAlaThrThr---GlnProTyr-----ThrTrpSerAsnValAla 58
Db 372857 CGGCGCTTCGGCAACTTCGGGGGCGCGCGCTCGGCTTTATGAACCTGGTCTCC-ACAACG 372799
Qy 59 IleGlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAla-----ProGly 76
Db 372798 TCGGGAATTCGGGCTTCCTC-----AACGTCGGCGCGCTGGGATCGGGT 372754
Qy 77 IleLeuTyrValArgThrAspIleGlyMetTyrArg-TipAspAlaAlaAsnGlyAr 96
Db 372753 GTGGGANTGTGGGCAACACCATCTCGGTATCTACACGTGGCAGCTCGGACCTCTCG 372694
Qy 96 gTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyrAsnGlyValValSe 116
Db 372693 ACGCGCGCGTAACTCCG-----GGTTGGCAA----- 372666
Qy 116 rIleAlaAlaAspProle-----AsnThrAsnLysVal----- 127
Db 372665 -----ATATCGGAACCAATATTCGCGGCTCTCGCGGACGCGCGGTACTCGGCTATT 372610
Qy 128 -----TrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIle 146
Db 372609 AACTGGGCTGGCCMACCGGCAACCTCAACGTGGGCTTCGCAAGTCTCGGCGGCTTT 372550
Qy 146 uArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGI 166
Db 372549 AACTTCG----- 372543
Qy 166 yAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnI 186
Db 372542 -----GCGGCGCCACCATCGGCCCAACAAC 372517
Qy 186 eLeuTyrPheGlyAlaPro---SerGlyLysGlyLeuTrpArgSer----- 200
Db 372516 GTCGGATCGGGAACACCGGAATCTTCGATGTGCGGCTTCGCGAACCCTGGGCACTCAAC 372457
Qy 201 -----ThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspVa 216
Db 372456 ATCGGCTTCGGAACCTTCGGCAGCAACCTCGGCTTCGGCAACTCG-----GC 372406
Qy 216 lGlyThrTrpIleAlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyVa 236
Db 372405 AGCTACAACATCGGCTTCGGCAACGTCGGCAACG-----AC 372370
Qy 236 lValTrpValAlaPheAspLysSerSer----- 245
Db 372369 AATCTGGGTTTCGCTAACCGCGGCGCGGCAACATCGGCTTCGGAACACCGGCAACAAC 372310
Qy 246 -----SerSerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAs 264
Db 372309 AATGTCGGCTTTGGGAACACGGGCGCAACAATGTGCGCATCGGGCTCACGGGCAACCGGA 372250
Qy 264 nProValPheTrpSerArgAspGlyGlyAla----- 274
Db 372249 C-----AGATCGGTTTCGGCAGCTTCAACTCGGCGGCAACATCGGC 372205
Qy 275 -----ThrTrpGlnAlaValProGlyAlaProThrGlyPh 286
Db 372204 CTGTTCACTCGGCGAGCAACAACATCGGATCTTCAATTCGGGCAAGCGCAACTTCGGC 372145
Qy 286 eIleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAs 306
Db 372144 -ATCGAACAATCGGGCAGCTTCAAC-----ACTGGCATCGGNA 372107
Qy 306 nThrGlyGlyProTyrAspGly-----SerSerGlyAspValTrpLysPheSerValTh 324
Db 372106 CACGGCAACCAATACCGGCTATTCAACTTCGGGCGAC-----GTCAA 372062

Qy 324 rSerGlyThrTrp---ThrArgIle-----SerProValProSerThrAspThrAlaAsnA 342
Db 372061 CACGGGCGCTTCAACCCGGGACGCTTCAACACCGGTAGCTTCAACACCGGCGAGCTTCAA 372002
Qy 342 spTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValA 362
Db 372001 CACCGGTGGCT---TCAATCGGGAATACCAACACCGGCTACCTCAACATTTGGCAACTA 371945
Qy 362 laThrGlnIleSerTrpTrpProAspThrIleIle-----PheArgSerThrAspGlyG 380
Db 371944 CAACACCGGATCGCCCAACACCGGCGAGTTGACACCGGGGCTTTCATCACCGGAACATA 371885
Qy 380 lyAlaThrTrp-----ThrArgIleTrpAspTrpThrSerTyrProA 394
Db 371884 CAGCAACGGGTGTTCTTAAGCGGCGATTACACGGGCTGG----- 371844
Qy 394 snArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrpIleuThrPheGlyValG 414
Db 371843 -----TCGGCTCAACCTGGTGATCGATATGCTCT 371813
Qy 414 lnPro-----AsnProProValProSerP 422
Db 371812 CCCCATAGCTCGGCGTGAATATTCATCGATATCCGATCACCGCTCGGCGGCA 371753
Qy 422 roLysLeuGlyTyr----- 426
Db 371752 CATCACCTTATGGCGCTCAGATTCCGCCACCGCGGATATGCTCTTCTGCTCAATAGC 371693
Qy 427 -----MetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyr--Gl 444
Db 371692 GGGCCAGGAGCCACCTTGGCCCCATTACCATTCGGAACATCACGGTTGTGGCCCCAC 371633
Qy 444 yThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlnI 464
Db 371632 GAGCACAGTCGCATAGAGGCGCGAATACCGCATACCATTAACCTGGCGGTGGCGCCAT 371573
Qy 464 eHisIle-----AlaProMetValLysGlyLeuGluThrAlaValAsn-- 479
Db 371572 TAGATCCCGCTCATCATGATATCCGCGCGCGCCAGGTTTCGGAACCTCGACCAACACC 371513
Qy 479 ----- 479
Db 371512 GTCGTCAGGTTTCTCAATACCGCGCGCGCGCTCGGGCTTCGGCAACTTCGGCGG 371453
Qy 480 -----AspLeuIleSerProSerGlyAlaProLeuIleSerAl 493
Db 371452 CGCCAAATTCGGGCTTTTGGAAACCTGGCTTCGCGACCTCGGGGCGCTCGGGCTCTCAA 371393
Qy 493 aLeuGlyAspLeuGly-----GlyPheThrHi 502
Db 371392 CGTCGGCGCTTCGGGATCAGGTCTCGGCAACGTGGGCAACCCGCTCTCGGGGTTCACAA 371333
Qy 502 sAlaAspValThrAlaValProSerThrIlePheThrSerProValPheThrGlyTh 522
Db 371332 CACGAGCATCGGACCTCGGAGCGCGGCTTCAATTCAGGCTTCGGCAACATCAGCAC 371273
Qy 522 rSerVal----- 524
Db 371272 CAGTATCGCGGCTGTCTGCGGACGACGCGGCAACCATGTCTCTCAACCTGGGCTTGCG 371213
Qy 525 -AspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSerSe 544
Db 371212 AAACCCAGGCACTCAACCTCGGCATT-----GC 371183
Qy 544 rGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrp----- 561
Db 371182 AAACCTCGGCACTACAACTCGGCTTTGCAACCTCGGAGCGGCAACTTCGGGCGCGC 371123
Qy 562 -----PheGlnGlySerGluProGlyGlyValThr----- 572
Db 371122 CAATATCGGTGGCAACAACATCGCGGCGGCAACACCGGGAATATTCGACATCGGTTTGGC 371063

```
QY 573 -----GlyGly-ThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaProG 590
Db 371062 AAATCTGGCAGTACATACATCGCTTCGGAACCTTCGGCATGACAACTGGGCTTC-G 371004
QY 590 lyAspProGlyGlnProValValTyAlaValGlyPheGlyAsnSerTrpAlaAlaSerG 610
Db 371003 GAAACCTCGGCAGC-----TACACGTCTGGCTTCGGAACCTTCGGCATGACAACTTCGG 370953
QY 610 InGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLySerPheTyra 630
Db 370952 TGGGCTTC-----GCCAACACCGGC----- 370933
QY 630 laLeuSerAsnGlyThrPheTyraArgSerThrAspGlyGlyValThrPheGlnProVala 650
Db 370932 -----AGCAACAATATCGGGTTCGCAACACCGGCAGC-----AACAAATATCG 370890
QY 650 laAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaValProGlyLyysG 670
Db 370889 GCATTGGGCTCACGGGGGACGGCAGATCGGG-----TTCGGCTCCCTGAATCTGGCA 370836
QY 670 luGlyAspLeuTrpLeu-----AlaAlaSerSerGlyLeuTyHisSerThrA 686
Db 370835 GCGGAACATCGGCTTTGTTCAACTCGGGCAGCGGAACATCGGCTTTTTCACATCGGGCA 370776
QY 686 snGly-----GlySerSerTrpS 692
Db 370775 ACGGAACGTTGGCATCGGCACACACCGGCACCAACTTCGGGCTTGGAAACACCGGCA 370716
QY 692 exAlaIleThrGlyVal-----SerSerAlaValAsnValGlyPheGlyLyysSerAla- 709
Db 370715 GCACCAACACCGGCTTCTTCAACTCGGGCAGCGTCAATACCGGTATCGGCAACACCGGCA 370656
QY 710 -----ProGlySerSerTyPrAlaValPheValValGlyT 722
Db 370655 GCTTCAACACCGCAGCTTCAATCCGGCGCATTCCAACACCGGGATTCACCCAGGCA 370596
QY 722 hrIleGlyGlyValThrGlyAlaTyArgSerAspAspCysGlyThrTrpValLeuI 742
Db 370595 GCTACAAC-----ACGGGACTCGGAACACACCGCGCATGTGACAC----- 370555
QY 742 leAsnAspAspGlnHisGlnTyGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaA 762
Db 370554 -----GGCGCTTCATCTCCGGCAGCTACAGCA 370527
QY 762 snLeuArgArgValTyIleGlyThrAsnGlyArgGlyIleValTyGlyAspIleGlyG 782
Db 370526 AC-----GGGTCTTGTGGAGTGGAAATATATCAGGGGCTCATTTGGCT 370485
QY 782 lyAlaProSerGlySerProSer-----ProSerValSerProSerAlaSerProS 799
Db 370484 T-GCAGCGCGCGTAGCGATTCCGGAATCGCCCTAACCTTTG-GCGTCGACATCCGAT 370427
QY 799 erLeuSerProSerProSer-----ProSerSerSerPro----- 810
Db 370426 ACATATACCATCAATCAATCAACCGCGGGTGTGTCACCCCTCCAGCGGCTTCAGCATCGTAGC 370367
QY 811 -----SerProSerProSerProSerProSerProSerProSerProSerP 826
Db 370366 TGCCGGAATAATATATGACTTACCCCATCATATCCGACCATCAATATCACTTGGCC 370307
QY 826 ro-----SerProSerProSerProSerProSerProSerProSerProSerAlas 843
Db 370306 CACGGCGCGCATCACCGTGGCGGACCCA-----CCACCTCGATCGGTATACCG----- 370257
QY 843 exProSerProSerProSerProSerProSerProSerProSerProSerP 860
Db 370256 --CCAGCGCGGTATCGGCTCCATCATCATCCGATCATCGACATTCCTCCGCGCATCGGG 370199
QY 860 roSerProThrProSerSerSerProValSerGlyGlyValLyysValGlnTyLyysAsnA 880
Db 370198 CTTTCGGCAACTCG-ACCACATAGTCGCGTCCGGGC-----TTCTTCAACT 370155
QY 880 sn-----AspSerAlaProGlyAspAsnGlnIleLyysProGlyLeuGln----- 894
```

```
Db 370154 CCGAGCGGCAGCGCGCTCGGGCTTTTGAACGTGGTCCGCGGCCTCAGGATTTCCG 370095
QY 895 -----ValValAsnThrGly----- 899
Db 370094 GTTATCTCAATGTCGGTGGCTGGGTGGGTGACTAACGTGGGTACACCGCTCTCGG 370035
QY 900 -----SerSerSerValAspLeuSerThrValThrValArgTyTrpPheThrA 916
Db 370034 GTTTCTACAACGCGAGCGGTGACCTCTGTGACGCGCGGCTTTTCCGCTCTCATGC 369975
QY 916 xAspGlyGlySerSerThrLeuValTyAsnCysAspTrpAlaAlaIleGlyCysGlyA 936
Db 369974 GCGAC-----GGTATGGGACCATGACTCTGAACCTTGGGCTGGCAACCTGGGCAACAATA 369918
QY 936 snIleArgAlaSerPheGlySer 943
Db 369917 AC-----GCCGGCTTCGGCAAC 369901
RESULT 15
US-09-252-991A-13873/c
; Sequence 13873, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 13873
; LENGTH: 3129
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13873
Alignment Scores:
Pred. No.: 1.73e-05 Length: 3129
Score: 280.00 Matches: 245
Percent Similarity: 33.12% Conservative: 114
Best Local Similarity: 22.60% Mismatches: 396
Query Match: 5.45% Indels: 233
DB: 4 Gaps: 61
US-09-917-376-1 (1-957) x US-09-252-991A-13873 (1-3129)
QY 36 ProfileAla-----IleThrAlaSerProAlaHisAlaAlaThrThr 49
Db 3113 CCGATCGCCAAACGGTACTGTGGTCAACGTGGCGGAGGAGCGCCGCGGTAAACAGCAGC 3054
QY 50 GlnProTyThrTrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleVal 69
Db 3053 CCGCGCGGCGAGC-----GTGACC 3036
QY 70 PheAsnGluGlyAlaProGlyIleLeuTyValArgThrAspIleGlyGlyMetTyArg 89
Db 3035 GTCGATTCAGCGCGCGCGCGCGCGGTGATCAACCGGAGCAACGCGGTCTCATCAGC 2976
QY 90 TrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsn----- 107
Db 2975 GGCACCGCGGAGCGCGGTGCCCGTACCTTACCGGAT---GCCGCGGCGCAACCGGATA 2919
QY 108 -----AsnTrpGlyTy----- 111
Db 2918 GGGCAGGTACCGCGGAGCGGCGGCACTGGAGCTTACCGCGGCGGCGCGCGGCGG 2859
QY 112 AsnGly-----ValValSerIleAlaAlaAspProfile---AsnThrAsnLysValTrp 128
Db 112 AsnGly-----ValValSerIleAlaAlaAspProfile---AsnThrAsnLysValTrp 128
```

2858 AACGGCAGCGTGATCGTCGCCACGCGCCACCGACCGGCAATACCGGCCCGCAGGCC 2799
Qy
129 Ala-----AlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAla 144
Db
2798 GCCACCAAGCGGTGGACGGGTGGCGCCGCGCGCGGTGATCGATCCGAGCAACGGCAGC 2739
Qy
145 IleLeu----- 146
Db
2738 ACCATCAGCGGCACCGCGGAGCGCGGCCCAAGGTGATCTCTACCGACGGCAACGGCAAC 2679
Qy
147 -----ArgSerSerAspGlnGlyAlaThrTrpGlnIleThrPro---LeuPro 161
Db
2678 CCGATCGCGGAAACCAACCGCGGAGCGGCGGCACTGGACCTTCAACCGCCCGCAGCGCG 2619
Qy
162 PheLysLeuGly-----GlyAsnMetProGly 170
Db
2618 CTGGCCAAACGGCAGCGGTGGTCAACCGCGTGGCCCGCAGGACCGTGGGGCAATACCGGCCCG 2559
Qy
171 ArgGlyMetGlyGluArgLeuAlaValAspProAsn-----Asn 183
Db
2558 CAGGGCAGCACTACCGTGGAGCGCGTGGCGCGGAACACGCGTGTGGTCAATCCGAGCAAC 2499
Qy
184 AsnAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeu----- 197
Db
2498 GGCAACCTGCTCAACGGTACCGCGGAGCGCGGAGCAACCGTGTGACCTTTGACCGACGGCAAC 2439
Qy
198 -----TrpArgSerThrAspSerGlyAlaThrTrp----- 207
Db
2438 GGCAAACCGGATCGGCCAGACCGCGCGGATGGCGAGCGGCACTGGAGCTTCAACCGCCGCGC 2379
Qy
208 SerGlnMetThrAsnPheProAspValGlyThrTrpIle---AlaAsnProThrAspThr 226
Db
2378 TCGCAACTACCCAAC-----GGCAACCGTGTCTCAACGTGACCGCGGAGCAGCC 2331
Qy
227 ThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSer 246
Db
2330 GCCGGCAATACCAAC-----GCTCCCGCTACCAACGAGCGGTGGATCTCTCGTCCG 2280
Qy
247 SerLeuGlyGln-----AlaSerLysThrIlePheValGlyValAlaAspProAsn 263
Db
2279 TCGATCCGAGGTGGATCCGAGCAACGGTTCGGTGTATCAGCGCAACCGCGGAGCAGCCGCGC 2220
Qy
264 AsnProValPhe-----TrpSerArgAspGly 272
Db
2219 AACACCATCATCATCACCAGTGGCAACCGCAACCGGATTCGGCCAGGTCAACCGCGCAGCGC 2160
Qy
273 GlyAlaThrTrpGlnAlaValProGlyAlaPro----- 283
Db
2159 AGCGTAACTGGTCTTCACTCCAGGCAATCCCGTGGCGGATGGCAGCGGTGTCACACGTG 2100
Qy
284 -----ThrGlyPhe 286
Db
2099 GTGGCGGCGCAGCCCAAGCAATGTCAGACGTGCGCGCGGTGATCACTGTGGATGGCGGTG 2040
Qy
287 IleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsn 306
Db
2039 GCCCGCGCGCGCGGTGATCGATCCAGCAACCGCACCGGCAAGATGACGGGTACCGCGGAG 1980
Qy
307 ThrGlyGlyProTyr-----AspGlySerSerGlyAspValTrpLysPheSerVal 323
Db
1979 GCCGGCGCAGCGGTGATCTCTACCGATGGCGGGGCAACCGGATCGCGCCAGCCACCGCC 1920
Qy
324 Thr---SerGlyThrTrpThrArgIleSerProValPro----- 335
Db
1919 GACGGCAGCGGCACTGGAGCTTCAACCGCGGCAACCGCGTGGCCAAACGGCACCGCGTATC 1860
Qy
336 -----SerThrAspThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAsp 352
Db
1859 AACCCCGTCCGAGGACCGCGCGCGCAATACAGCGGTCCGGCCACCGTCAACCGTGTAT 1800
Qy
353 ArgGlnHisPro-----AsnThrIleMetValAlaThrGln 364
Db
1799 GCCATCGCCCGCGCGCGGTGATCAATCCGAGCAACCGAGTGTGTCATCAGCGGTACG 1740

365 IleSerTrpTrpProAspThrIleIlePheArgSerThrAspGlyGly----- 380
Db
1739 GCGGAAGCGCGGCGCACGGTGTATCTC-----ACGACGGCAACGGCAACCGCATC 1689
Qy
381 -----AlaThrTrpThrArgIleTrpAspTrpThr---SerTyrProAsnArg----- 395
Db
1688 GCCAGGTGTCACCGCGCAGCGGCAAGTGGGTTCACGCCCGCCACCGCTGTGGCC 1629
Qy
396 -----SerLeuArgTyrValLeuAspIleSerAlaGluProTrpLeuThrPh 411
Db
1628 AATGGCAGCGGTGATCAATGCGCTGG-----CCAGGACGCGCGCGCACCAAC 1581
Qy
411 eGlyValGlnProAsnProProValProSerProLysLeuGlyTrpMetAspGlyAlaMe 431
Db
1580 AGCAGTC-----CCACCAGGGCCA---CCGTCGACTCGCTGGCGGCCACGA---GCCCC 1534
Qy
431 tAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThr-----GlyAlaTh 448
Db
1533 GGTGATCGATCCGAGCAACGGTAGC---GTATCGCGGTATCCGCGAGGCTGTGGCCAC 1477
Qy
448 rLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAlaPr 468
Db
1476 GGTGATCTCTACC-----GACGGCAACGGCAAC-----CC 1447
Qy
468 oMetValLysGlyLeuGluGluThrAlaValAsnAspLeuIleSerProProSerGlyAl 488
Db
1446 GATCGGCCAGTCTACCGCGGATGGCAGCGGCACTGGAGTTCAGCGCC-----GGCAC 1393
Qy
488 aProLeu-----IleSerAlaLeuGly---AspLeuGlyGlyPheThrHi 502
Db
1392 GCCGTGTCCAATGCGACGGTGTCAATCGGTGGCCAGGACGCTGCGGCAACACACAG 1333
Qy
502 sAlaAspVal---ThrAlaValProSerThrIlePheThrSerProValPhe-----Th 519
Db
1332 CGGCCCGGCCAGCACCGTGGACTCGGTGGCGCGCGCGCGGTGATCGACCCCGAG 1273
Qy
519 rThrGlyThrSerVal---AspTyrAlaGluLeuAsnProSerIleIleValArgAlaGl 538
Db
1272 CAACGGCAGCGTGTATCGCGGTACCGGAGACCGCGTGGCGGTGATCTCACCAGTGG 1213
Qy
538 ySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGly 558
Db
1212 CGCGCGCAACCGATCGGCCAGGCC-----ACCGCGATGGCAG 1174
Qy
558 yLysAsnTrp-----PheGl 563
Db
1173 CGGCAACTGGAGTTCACCCCGGCGCACCGGTGGCCAAACGGCACGGTGTATCAATGCGGT 1114
Qy
563 nGlySerGluProGlyValThrThrGlyGlyThr----- 575
Db
1113 GGCCAGGATCCGGCGGCAATACCGCGCGCCGACAGCACCGGTGGAGCGCGGTGGC 1054
Qy
576 -----ValAlaAlaSerAlaAspGl 582
Db
1053 CCGCGCCACCCCGGTGGTCAACCCGAGCAACGGCAGCGTATCGCGGTACCGCGGAACG 994
Qy
582 ySerArgPheValTrpAlaProGlyAspProGlyGlnProVal-----ValTyrAlaVa 600
Db
993 CGCGCCACCGGTGATCTCTACCGCAGCGGCGGCAACCGGATCGCGCCAGGTCAACCGCGCA 934
Qy
600 lGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValPro---AlaAsnAlaGlnIleAr 619
Db
933 CGGACCGGGCAAC---TGGAGCTTCAACGCCCGGCAACCGCTGGCCAAACGGTCTCGGTAT 877
Qy
619 gSer-----As 621
Db
876 CAATGCGTGGCCAGGACCGCGGCAACACCGAGCGCGCGCCAGCACCGACCGGTGA 817
Qy
621 pArgValAsnProLysThr---PheTyrAlaLeuSerAsnGlyThrPheTyrArgSer--- 639
Db
816 CTGCGTAGCCCGGCCACCCCGGTGCTCGATCCGAGCAACCGTACGGTGTATCAGCGGTAC 757

```
Qy 640 -----ThrAspGlyGlyValThrPheGlnProValAl 650
Db 756 CGCGAAGCGGGGCCACCGGTGATCTCACCAGCGCGGC-----GGCAACCCGGATACG 703
Qy 650 aAlaGlyLeu-ProSerSerGlyAlaValMetPheHisAlaValProGlyLysG 670
Db 702 GCAGGCCACCGCGATGCGCGGCACTGGAGCTTCACCTCCGGGCACA-CCGCTGACCA 644
Qy 670 luGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyHisSerThrAsnGly-----G 688
Db 643 ACGGCAGGTGATCATATGCGGTGGCCAGGACCGCGCGGCAACACACGAGCGTCCGGTCA 584
Qy 688 lySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGlyPheGlyLysS 708
Db 583 GCACCACAGTGGACCGGTGGCCCGCGCCACCGCGGTGATCGACCCGAGCAATGGTGTC 524
Qy 708 erAlaProGlySerSerTyProAlaValPheValValGlyThr-----IleGlyGlyV 726
Db 523 AACTCAGCGGCACCGCGGAACCGGGTCCGGGTGATCTCACCAGTGGCAATGGCAACC 464
Qy 726 alThrGlyAlaTyArgSerAspAspCysGlyThrThrTrpValLeuIleAsnAspAspG 746
Db 463 CGATCGGCCAGACCTCGCCGACGGTAGCGGC---AACTGGACCTTCACACCGGGCACGC 407
Qy 746 lnHisGlnTyArgAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgV 766
Db 406 CGCTGGCCAAACGGCACGGTGGTCAACCGCTGGGCCAGGACCCGCGCGGC----- 357
Qy 766 alTyIleGlyThrAsnGlyArgGlyIleValTyGlyAspIleGlyGlyAla---ProS 785
Db 356 -----AATACAGCGGTCC-GGCCAGCACCCAGTGGATACGGTGGCGCGCGCCACG 306
Qy 785 erGlySerProSerProSerValSerProSerAlaSerProSerLeuSerProSerProS 805
Db 305 CCGGTGATCA-ATGCCA-----GCCACGGCAGCGGTGATCACCAGCACCG- 263
Qy 805 exProSerSerProSerProSerProSerProSerProSerProSerProSerP 824
Db 262 --CGAGGTGGCGCCAAAGTATCTCTCACCGACGGCAACGGCAACCGGATCGCGGAGAC 205
Qy 824 roSerPro-----SerProSerProSerProSerProSerProSerProSerProS 841
Db 204 CACCGCGCAGCGCAGTGGCAACTGGACCTTCACCCCGCGCACGCGCTGGCCCAACGGTAC 145
Qy 841 er---AlaSerProSerProSerProSerProSerProSerProSerProSerP 860
Db 144 GGTGATCAACCGCTCGCCGAAG-----ACGCCGGGGCAACGCCAGCGGTCCGGC 94
Qy 860 roSerPro 862
Db 93 CAGCACCA 86
```

Search completed: May 14, 2004, 17:05:37
Job time : 8674 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 14, 2004, 14:54:31 ; Search time 941 Seconds
(without alignments)
4615.210 Million cell updates/sec

Title: US-09-917-376-1

Perfect score: 5135

Sequence: 1 MORSSEIRLTMRSRLVSL.....RASFGSVNPATPTADTYLOX 957

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2947324 seqs, 2269024515 residues

Total number of hits satisfying chosen parameters: 5894648

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DRV=xlh
-Q=/cgn2_1/USPTO.spool/US09917376/runat_14052004_085258_4727/app.query.fasta_1.1095
-DB=Published Applications NA -QFMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62
-TRANS=human40.cdi -LIST=45 -DLOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09917376@cgn_1_621@runat_14052004_085258_4727
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=130 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
- 17: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 18: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 19: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
------------	-------	-------	--------	----	----	-------------

ALIGNMENTS

RESULT 1

US-09-917-376-2

; Sequence 2, Application US/09917376

; Publication No. US20040038334A1

; GENERAL INFORMATION:

; APPLICANT: DING, SHI-YOU

; APPLICANT: ADNEY, WILLIAM S.

; APPLICANT: VINZANT, TODD B.

; APPLICANT: HIMMEL, MICHAEL E.

; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS

; FILE REFERENCE: 40197.4US01

; CURRENT APPLICATION NUMBER: US/09/917,376

; CURRENT FILING DATE: 2001-07-28

; NUMBER OF SEQ ID NOS: 7

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2

; LENGTH: 2869

; TYPE: DNA

; ORGANISM: Acidothermus cellulolyticus

; FEATURE:

; NAME/KEY: modified_base

; LOCATION: (2869)

1	5134	100.0	2869	13	US-09-917-376-2	Sequence 2, Appli
2	5134	100.0	2869	15	US-10-155-400-2	Sequence 2, Appli
3	2487	48.3	9025608	15	US-10-156-761-1	Sequence 1, Appli
4	2478	48.3	2646	15	US-10-156-761-1845	Sequence 1845, Ap
5	1631.5	31.8	2517	15	US-10-026-994-4	Sequence 4, Appli
6	1631.5	31.8	2710	15	US-10-026-994-1	Sequence 1, Appli
c	7	1570	30.6	9025608	15	US-10-156-761-1
8	1167.5	22.7	3668	10	US-09-927-827-21	Sequence 2561, Ap
9	1167.5	22.7	2646	13	US-10-395-241-11	Sequence 11, Appl
10	1125.5	21.9	2646	13	US-10-395-241-17	Sequence 17, Appl
11	1124	21.8	2481	13	US-10-395-241-13	Sequence 13, Appl
12	1117	21.8	2367	13	US-10-420-191-1	Sequence 1, Appli
13	940	18.3	5698	10	US-09-917-384-2	Sequence 2, Appli
14	823	16.0	3687	10	US-09-917-383-2	Sequence 2, Appli
15	823	16.0	3287	10	US-09-917-378-2	Sequence 2, Appli
16	702	13.7	2223	15	US-10-156-761-550	Sequence 550, App
17	361	7.0	2223	15	US-10-087-192-289	Sequence 289, App
c	18	345	6.7	18596	9	US-09-880-107-1590
c	19	345	6.7	18596	9	US-09-967-768A-119
c	20	345	6.7	18596	9	US-09-954-531-124
c	21	345	6.7	18596	9	US-09-954-531-348
22	330.5	6.4	7407	13	US-10-282-122A-30151	Sequence 30151, A
23	330.5	6.4	7407	15	US-10-246-330-3	Sequence 3, Appli
24	299.5	5.8	210528	13	US-10-087-192-289	Sequence 289, App
25	297	5.8	32591	16	US-10-085-117-187	Sequence 187, App
26	296.5	5.8	4767	10	US-09-940-316B-28	Sequence 28, Appl
27	296	5.8	96602	16	US-10-085-117-61	Sequence 61, Appl
28	293	5.7	4818	10	US-09-940-316B-32	Sequence 32, Appl
c	29	292.5	5.7	185548	15	US-10-175-523-62
c	30	289	5.6	324	13	US-10-282-122A-13296
31	288.5	5.6	4571	10	US-09-940-316B-18	Sequence 18, Appl
32	288	5.6	288	16	US-10-437-708-111	Sequence 111, App
33	288	5.6	288	16	US-10-395-402-111	Sequence 111, App
34	285.5	5.6	5184	15	US-10-260-715-7	Sequence 7, Appli
35	282.5	5.5	9903	13	US-10-282-122A-28185	Sequence 28185, A
36	282	5.5	85692	17	US-10-461-194-1	Sequence 1, Appli
37	281.5	5.5	432	9	US-09-864-761-32943	Sequence 32943, A
38	281.5	5.5	518	9	US-09-864-761-16460	Sequence 16460, A
39	280	5.5	86941	17	US-10-461-194-2	Sequence 2, Appli
40	279	5.4	4547	10	US-09-940-316B-22	Sequence 22, Appl
41	275	5.4	666	15	US-10-029-386-25948	Sequence 25948, A
c	42	275	5.4	38918	15	US-10-017-161-2049
c	43	275	5.4	38918	16	US-10-292-798-1695
c	44	274.5	5.3	1168	15	US-10-017-161-2179
c	45	274.5	5.3	1168	16	US-10-292-798-1825

OTHER INFORMATION: a, c, t, g, other or unknown
US-09-917-376-2

Alignment Scores:

Pred. No.: 0 Length: 2869
Score: 5134.00 Matches: 956
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.98% Indels: 0
DB: 13 Gaps: 0

US-09-917-376-1 (1-957) x US-09-917-376-2 (1-2869)

```
QY 1 MetAspArgSerGluAsnIleArgLeuThrMetArgSerArgArgLeuValSerLeuLeu 20
DB 1 ATGGATCGTTCGGAGAACATCCGCTCTGACTATGAGATCAGACGATTGGTATCACTGCTC 60
QY 21 AlaAlaThrAlaSerPheAlaValAlaAlaAlaLeuGlyValLeuProIleAlaIleThr 40
DB 61 GCGGCCACTGCGTCTGTCGGCGGCGCGCTCTGGGAGTTCTGCCCATTCGGGATAACG 120
QY 41 AlaSerProAlaHisAlaAlaThrThrGlnProTyrThrTrpSerAsnValAlaIleGly 60
DB 121 GCTTCTCTCGGCACGCGGCGGAGACTACGCCGTACACCTGGAGCAACGTGGCGATCCGG 180
QY 61 GlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrVal 80
DB 181 GCGCGCGGCTTCTCGACGGGATCGTCTTCAATGAAGGTGCACCGGGAATCTGTAGTGT 240
QY 81 ArgThrAspIleGlyGlyMetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeu 100
DB 241 CGGACGGACATCGGGGGGATGTATCGATGGATGCGCCCAACGGGGGTGGATCCCTCTT 300
QY 101 LeuAspTrpValGlyTrpAsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAsp 120
DB 301 CTGGATTGGGTGGATGGAAACAATTGGGGGTACAACGGCGCTGTACAGCATTCGGCGAGC 360
QY 121 ProIleAsnThrAsnIleValTrpAlaAlaValAlaGlyMetTyrThrAsnSerTrpAspPro 140
DB 361 CCGATCAATACTAACCAAGGTATGGCGCGCGCTGGGAATGTACACCAACAGCTGGACCCA 420
QY 141 AsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeu 160
DB 421 AACGACGGAGCATTTCTCGCTCGTCTGATCAGCGCGCAACGTGGCAATAAAGCGCCCTG 480
QY 161 ProPheIleLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAsp 180
DB 481 CCGTTCAAGCTTGGCGGCAACATGCCCGGCGGTGGAAATGGCGGAGCGGCTTGGGGTGGAT 540
QY 181 ProAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyIleGlyLeuTrpArgSer 200
DB 541 CCAAAACAATGACAAACATTCGTATTTTCGGCGCGCCCGCGGCGGAAAGGCTCTGGAGAAC 600
QY 201 ThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyrIle 220
DB 601 ACAGATTCCGCGCGGACCTGCTCCAGATGACGAACCTTCCGGAGCTAGGACGATACATT 660
QY 221 AlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAla 240
DB 661 GCAAAATCCCACTGACACGACCGGCTATCAGAGCGGATATTCAGGCGGTCTGGGTGCT 720
QY 241 PheAspIleSerSerSerLeuGlyGlnAlaSerIleThrIlePheValGlyValAla 260
DB 721 TTCGAAAGTCTTCGTATCGCTCGGCGCAAGCAGATGAAGCAATTTTGTGGCGGTGGCG 780
QY 261 AspProAsnAsnProValPheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValPro 280
DB 781 GATCCCAATTAATCCGCTCTTCGAGCAGACGCGCGCGCGAGCTGGCAGGCGGTCCG 840
QY 281 GlyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsnHisValLeu 300
DB 841 GGTGCGCGGACGGCTTCATCCGCGCAAGGGGCTCTTTGACCCGGTCAACCAACGCTGCTC 900
```

```
QY 301 TyrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspValTrpIlys 320
DB 901 TATATTTCACCAACAGCAATACGGTGGTCCGTATACGGGAGCTCCGGGACGCTCGGAAA 960
QY 321 PheSerValThrSerGlyThrTrpThrArgIleSerProValProSerThrAspThrAla 340
DB 961 TTCTCGGTGACCTCCGGGACATGACCGGAATCAGCCCGGTACCTTCGACGGACACGGCC 1020
QY 341 AsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMet 360
DB 1021 AACGACTACTTTGGTTACAGCGGCTCACTATCGACCGCCAGCACCCGCAACACGATAATG 1080
QY 361 ValAlaThrGlnIleSerTrpTrpProAspThrIleIlePheArgSerThrAspGlyGly 380
DB 1081 GTGCGCAACCCAGATATCGTGGTGGCGGACACCAATAATCTTTTCGGAGCACCGACGCGGT 1140
QY 381 AlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrVal 400
DB 1141 GCGACGTGGACGCGGATCTGGGATGGACGAGTTATCCCAATCGAAGCTTTCGATATGTG 1200
QY 401 LeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnProProValPro 420
DB 1201 CTTGACATTTTCGGGAGGACCTTGGCTGACCTTCGGGTACAGCCGAATCCTCCCGTACCC 1260
QY 421 SerProIysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArg 440
DB 1261 AGTCCGAAGCTCGGCTGGATGGATGAAGCGATGGCAATCGATCCGTTCAACTCTCATCGG 1320
QY 441 MetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrIleThrAspSer 460
DB 1321 ATGCTCTACCGAACAGCGCGGACGTTGTGACCAACAAATGATCTCAGGAAGTGGGACTCC 1380
QY 461 GlyGlyGlnIleHisIleAlaProMetValIysGlyLeuGluThrAlaValAsnAsp 480
DB 1381 GCGCGCCAGATTCAATATCGGCCCGATGGTCAAGGATGGAGGACGCGCGTAAACGAT 1440
QY 481 LeuIleSerProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPhe 500
DB 1441 CTCATCAGCGCGCGCTTGGCGCGCGCTCATCAGCGCTCTCGGAGACTCGCGCGCTTC 1500
QY 501 ThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProValPheThrThr 520
DB 1501 ACCACCGCCAGCTTACTTGGCGTGCCTACGACGATCTTCACGTCACCGGTGTTCAGACC 1560
QY 521 GlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPhe 540
DB 1561 GGCACACGCTGCACTATCGGAATTGAAATCCGTCGATCATCGTTCGGCTGGAGATTTC 1620
QY 541 AspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyIysAsn 560
DB 1621 GATCCATCGAGCCAAACCGAACGACAGGACGCTCGGTTCTCGACAGACGCGCGCAAGAAC 1680
QY 561 TrpPheGlnGlySerGluProGlyValThrThrGlyGlyThrValAlaAlaSerAla 580
DB 1681 TGGTTCCAAGCAGCAGAACTTGGCGGGGTGACGAGCGCGGACCGCTCGCCCATTCGGCC 1740
QY 581 AspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValTrpAlaVal 600
DB 1741 GACGGCTCTCGTTTCGTCTGGGCTCCCGCGCATCCCGGTACGCTGTGTGTACGAGTTC 1800
QY 601 GlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSer 620
DB 1801 GGATTTGGCAACTCTCGGCTGCTTCGCAAGGTGTTCGCGCAATGCCAGATCCGCTCA 1860
QY 621 AspArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThr 640
DB 1861 GACCGGTGAATCCAAAGACTTCTATGCCCTATCCAAATGGAACTTCTATCGAAGCAGC 1920
QY 641 AspGlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGly 660
DB 1921 GACGCGCGCGTGACATTCCAACCGGCTCGCGCGGCTCTTCGAGAGCGGCGGTCCGCTCGGT 1980
QY 661 ValMetPheHisAlaValProGlyLysGlyAspLeuTrpLeuAlaAlaSerSerGly 680
```

```
Db 1981 GTGATGTTCCACGGCGCTGGAAGAGAGCGGATCTGGGCTGCTGCATCAGCGGG 2040
Qy LeuTyrHisSerThrAsnGlySerSerTyrAlaIleThrGlyValSerSerAla 700
Db 2041 CTTTACCACTACCAATGGCGGAGAGTGGTCTGCATCAGCGGATCTCCGCG 2100
Qy ValAsnValGlyPheGlySerSerAlaProGlySerSerTyrProAlaValPheVal 720
Db 2101 GTGAACGTGGATTTGGTAAGTCTGCGCGCGGTCGTATACCCAGCGCTTTGTGCTC 2160
Qy GlyThrIleGlyValThrGlyAlaTyrArgSerAspAspCysGlyThrTyrVal 740
Db 2161 GGCACGATCGGAGCGGTATACGGGGCGGTACCGCTCCGACGACTGTGGAGCAGCTGGGTA 2220
Qy LeuIleAsnAspAspGlnHisGlnTyrGlyAsnTyrGlyGlnAlaIleThrGlyAspHis 760
Db 2221 CTGATCANTGATGACACGACCAATACGGAATTTGGGACAGCAATACCGGTTGACAC 2280
Qy AlaAsnLeuArgArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIle 780
Db 2281 GCGAATTTACGGCGGTGTACATAGGACAGACGCGCGTGAATTTGTATACGGGACATT 2340
Qy GlyAlaProSerGlySerProSerProSerValSerProSerAlaSerProSerLeu 800
Db 2341 GGTGGTGGCGCTCCGATCGCGCTCCGTGGTGGTCCGTGGTGGTCCGCGAGCGCTG 2400
Qy SerProSerProSerProSerProSerProSerProSerProSerProSerProSerPro 820
Db 2401 AGCCGAGCCGAGCCGAGCAGCTCCCATCGCCGTCGCGCTCCGCGAGCTCGAGTCCA 2460
Qy SerSerSerProSerProSerProSerProSerProSerProSerProSerProSerPro 840
Db 2461 TCCTGTCGCGCTCTCGTCCGCTACCATCGCCGAGTCCGTCTCGGTCTCCGTACCA 2520
Qy SerAlaSerProSerProSerProSerProSerProSerProSerProSerProSerPro 860
Db 2521 TCGCGCTCGCGAGCGCTCTCTGTCACCGAGCCCGCTTCCTCGTCCGCTCGCG 2580
Qy SerProThrProSerSerSerProValSerGlyValValValGlnValValAsnThrGlySer 880
Db 2581 AGCCCAACGCGCTCGTCCGCGCTGTCGCGTGGGTTGAGGTGAGTATAGAAATAAT 2640
Qy AspSerAlaProGlyAspAsnGlnIleLysProGlyLeuGlnValValAsnThrGlySer 900
Db 2641 GATTTCGGCGCGGTGATATAGATCAAGCCGGGTTTGACGGTGGTGAATACCGGGTCG 2700
Qy SerSerValAspLeuSerThrValThrValArgTyrTyrPheThrArgAspGlyGlySer 920
Db 2701 TCGTCCGTGGATTTGTGACGGTGACGGTGGGTACTGTTTACCCGGGATGGTGGCTCG 2760
Qy SerThrLeuValTyrAsnCysAspTyrAlaAlaIleGlyCysGlyAsnIleArgAlaSer 940
Db 2761 TCGACACTGGTGTACAACTGTGACTGGCGCGCATCGGGTGGGAATATCCGCGCTCG 2820
Qy PheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 956
Db 2821 TTGGCTCGGTGAACCGCGGACGCCGACGCGGACACCTACTCTGCAG 2868
```

RESULT 2

```
US-10-155-400-2
; Sequence 2, Application US/10155400
; Publication No. US2003010898A1
; GENERAL INFORMATION:
; APPLICANT: ADNEY, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS
; FILE REFERENCE: NREL 01-36A
; CURRENT APPLICATION NUMBER: US/10/155,400
; CURRENT FILING DATE: 2002-10-22
```

```
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2869
; TYPE: DNA
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:
; NAME/KEY: modified base
; LOCATION: (2869)
; OTHER INFORMATION: a, c, t, g, other or unknown
US-10-155-400-2
```

Alignment Scores:

```
Pred. No.: 0 Length: 2869
Score: 5134.00 Matches: 956
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.98% Indels: 0
DB: 15 Gaps: 0
```

US-09-917-376-1 (1-957) x US-10-155-400-2 (1-2869)

```
Qy 1 MetAspArgSerGluAsnIleArgLeuThrMetArgSerArgArgLeuValSerIleLeu 20
Db 1 ATGGATCGTTTCGGAGAACATCCGCTGCTGATATGAGATCAGACGATTTGGTATCAGTCTC 60
Qy 21 AlaAlaThrAlaSerPheAlaValAlaAlaLeuGlyValLeuProIleAlaIleThr 40
Db 61 GCCGCCACTGCGTGTTCGCCGTCGCCGCTCTGGGAGTTCTGCCCATCGGATAACG 120
Qy 41 AlaSerProAlaHisAlaAlaThrThrGlnProTyrThrTyrSerAsnValAlaIleGly 60
Db 121 GCTTCTCTCGCAGCGCGCGGAGCTCAGCGGTACACTGGAGCAACCTGGCGATCGGG 180
Qy 61 GlyGlyGlyPheValAspGlyIleValPheAsnGlyAlaProGlyIleLeuTyrVal 80
Db 181 GCGCGCGCTTTGTTCGACGGGATCGTCTTAATGAAGGTGCACCGGGAATTTCTGTACGTG 240
Qy 81 ArgThrAspIleGlyGlyMetTyrArgTyrAspAlaAlaAsnGlyArgTyrIleProLeu 100
Db 241 CGSAGCGGACATCGGGGGGATGTATCGATGGATGCCGCCAAGCGGGCGGTGGATCCCTCTT 300
Qy 101 LeuAspTyrValGlyTyrAsnAsnTyrGlyTyrAsnGlyValValSerIleAlaAlaAsp 120
Db 301 CTGATTTGGTGGATGAGAACATTTGGGGGTACACGCGGTCTGTACGATTTGGCGGAGAC 360
Qy 121 ProIleAsnThrAsnLysValTyrAlaAlaValGlyMetTyrThrAsnSerTyrAspPro 140
Db 361 CCGATCAATACTAACAAGGTATGGCGCGCTGCGGAATGTACACCAACAGCTGGGACCCA 420
Qy 141 AsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTyrGlnIleThrProLeu 160
Db 421 AACGACGAGCGATTTCTCCGCTCGTCTGATCAGGCGCAACGTGGCAATAACGCCCTCG 480
Qy 161 ProPheIleLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAsp 180
Db 481 CCGTTCAAGCTTGGCGGCAACATGCCCGGGCTGGATGGCGGAGCGGTCTTCGCGGTGAT 540
Qy 181 ProAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTyrArgSer 200
Db 541 CCNAACTGACACATTTCTGTTTCGCGCGCCCGGAGCGGCAAGGGCTCTGGAGAAGC 600
Qy 201 ThrAspSerGlyAlaThrTyrSerGlnMetThrAsnPheProAspValGlyThrTyrIle 220
Db 601 ACAGATTCGGCGGACCTGCTGCCAGATGACGAATTTTCGGAGCTAGGACGATACATT 660
Qy 221 AlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTyrValAla 240
Db 661 GCAATCCCTGACGACCGGCTATCAGAGCGGATATTCAGAGCGGTCTGTCTGGGTGCT 720
Qy 241 PheAspLysSerSerSerLeuGlyGlnAlaSerLysThrIlePheValGlyValAla 260
Db 721 TTCGACAGTCTTCGTCTCATCGCTCGGCGAGGAGTAGACCATTTTGTGGCGGTGGCG 780
```

QY 261 AspProAsnAsnProValPheThrSerArgAspGlyGlyAlaThrTrpGlnAlaValPro 280
DB 781 GATCCCAATAATCCGGTCTTCTGAGCAGAGACGGCGCGCGAGCTGGCAGGCGGTGCGC 840
QY 281 GlyAlaProThrGlyPheThrProHisLysGlyValPheAspProValAsnHisValLeu 300
DB 841 GGTGCCCGCCAGCGGCTTCAATCCCGCAAGAGGGCGTCTTTACCCGGTCAACACGAGTGC 900
QY 301 TyrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLys 320
DB 901 TATATTGCCACACAGCAATACGGGTGTCCTGATGACGGGAGCTCCGGCAGCTCTGGAAA 960
QY 321 PheSerValThrSerGlyThrTrpThrArgIleSerProValProSerThrAspThrAla 340
DB 961 TTCTCGGTGACCTCCGGGACATGGACGCAATCAGCCCGGTACTCTTCGACGGCACCGGCC 1020
QY 341 AsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMet 360
DB 1021 AACGACTACTTTGGTTACAGCGGCTCCTATCGACCGCGCAGCACCCGAAACAGATAATG 1080
QY 361 ValAlaThrGlnIleSerTrpTrpAspThrIleIlePheArgSerThrAspGlyGly 380
DB 1081 GTGGCAACCCAGATATCGTGGTGGCCGGAACCAATAATCTTTTCGAGCACCGAGCGCGT 1140
QY 381 AlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrVal 400
DB 1141 GCGACGTGGACGCGGATCTGGGATTGACGAGTTATCCCAATCGAAGCTTGGCATATGTG 1200
QY 401 LeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnProValPro 420
DB 1201 CTTTGACATTTCCGCGGAGCCTTGGCTGACCTTCGGCGTACAGCGGAATCTCTCCGTACCC 1260
QY 421 SerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArg 440
DB 1261 AGTCCGAAGCTCGGTGGATGGATGAAGCATGGAATCGATCCGTTCAACTCTGATCGG 1320
QY 441 MetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSer 460
DB 1321 ATGCTCTACGGACAGCGCGGACGTTGTACGCAACAATGATCTCAGCAAGTGGGACTCC 1380
QY 461 GlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluThrAlaValAsnAsp 480
DB 1381 GCGCGCCAGATTATATCGCGCGCATGGTCAAGGATTGGAGGAGACGGCGGTAAACGAT 1440
QY 481 LeuIleSerProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPhe 500
DB 1441 CTATCAGCCCGCGCTTGGCGCCCGCTCATCAGCGCTCTCGGAGACCTCGCGCGGCTTC 1500
QY 501 ThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProValPheThrThr 520
DB 1501 ACCCAGCGCAGCTTACTGCGTGCCATCGACGATCTTCAGTCAACCGGTGTTCCAGACC 1560
QY 521 GlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValAlaGlySerPhe 540
DB 1561 GGCACACGCGTACATATGCGGAATGAATCCGTCGATCATCGTTCGCGCTGGAAATTC 1620
QY 541 AspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsn 560
DB 1621 GATCCATCGAGCCCAACCGCAACGACGACGCTCGCGTCTTCGACAGACGGCGGCAAGAC 1680
QY 561 TrpPheGlnGlySerGluProGlyGlyValThrThrGlyGlyThrValAlaAlaSerAla 580
DB 1681 TGGTTTCAGGACGAGCACTTGGCGGGGTGACGACGGCGGACCGCTCGCGCATTCGCCC 1740
QY 581 AspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValTrpAlaVal 600
DB 1741 GACGGCTCTGTTCTGCTCGGCTCCCGCGATCCCGGTGAGCTGTGGTGTACGAGTC 1800
QY 601 GlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSer 620
DB 1801 GGATTTGGCAACTCTCGGTGCTTCGCAAGGTGTTCCCGCAATGCCAGATCCCGCTCA 1860

QY 621 AspArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThr 640
DB 1861 GACCGGTGATATCCAAAGACTTTCTATCCCTATCCATGGAACCTTTCTATCGAAGCAG 1920
QY 641 AspGlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGly 660
DB 1921 GACCGCGCGTGACATTTCCAAACCGTTCGCGCGCGGTCTTCGAGGACGGGTCCGTCCGT 1980
QY 661 ValMetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGly 680
DB 1981 GTCATGTTCCACGCGGTGCTTGGAAAGAGAGCGCATCTGTGGCTCGTCATCGAGCGGG 2040
QY 681 LeuTyrHisSerThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAla 700
DB 2041 CTTTACCACCTCAACCAATGGCGGAGCAGTTGGTCTGCAATCACCGCGGTATTCCTCCGCG 2100
QY 701 ValAsnValGlyPheGlyLysSerAlaProGlyLysSerTyrProAlaValPheValVal 720
DB 2101 GTGAACGTGGATTTGGTAACTCTGCGCCCGGTCTCATACCCAGCGCTCTTTGTCTGTC 2160
QY 721 GlyThrIleGlyGlyValThrGlyValAlaTyrArgSerAspCysGlyThrThrTrpVal 740
DB 2161 GGCACGATCGGAGCGGTTACGGGGCGGTACCGCTCCGACGACTGTGGGACGACCTGGGTA 2220
QY 741 LeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHis 760
DB 2221 CTGATCAATGATGACACGACCAATACGAAATTTGGGACAGCAATCACCGGTGACCAAC 2280
QY 761 AlaAsnLeuArgArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIle 780
DB 2281 GCGAAATTTACGGCGGGTGTACATAGGACAGAACGCGCGTGGAAATGTATACGGGACAT 2340
QY 781 GlyGlyAlaProSerGlySerProSerProSerValSerProSerAlaSerProSerLeu 800
DB 2341 GGTGGTGGCGCTCGGATCCGCGTCTCCGTGGGTGAGTCCGTCCGCTTCGCGGAGCTG 2400
QY 801 SerProSerProSerProSerProSerProSerProSerProSerProSerProSerPro 820
DB 2401 AGCCGAGCGCGAGCGCGAGCAGCTCGCATCGCGCTCGCGCTCGCGAGCTCGAGTCCA 2460
QY 821 SerSerSerProSerProSerProSerProSerProSerProSerProSerProSerPro 840
DB 2461 TCCTCGTGGCGCTTCGCGTCCCGTCCCATCGCGAGTCCGTCTCGGTCTCCGTCCCA 2520
QY 841 SerAlaSerProSerProSerProSerProSerProSerProSerProSerProSerPro 860
DB 2521 TCGCGCTCGCGACCGCGCTTCGTCCACGAGCCCGTCTTCGTCCACCGCTTCGTCCGCG 2580
QY 861 SerProThrProSerSerSerProValSerGlyValLysValGlnTyrLysAsnAsn 880
DB 2581 AGCCCAACGCGCTCGTTCGTCGCGGTGGGTGGGTGAGGTGAGGTGAGTAAATAAT 2640
QY 881 AspSerAlaProGlyAspAsnGlnIleLysProGlyLeuGlnValValAsnThrGlySer 900
DB 2641 GATTTCGCGCGCGGTGATTAATCAGATCAAGCGCGGTTCGAGTGGTGAATACCGGCTCG 2700
QY 901 SerSerValAspLeuSerThrValThrValArgTyrTrpPheThrArgAspGlyGlySer 920
DB 2701 TCGTGGTGGATTTGTGACCGGTGACCGTGGGTGAGTGGTTCACCGGAGATGGTGGCTCG 2760
QY 921 SerThrLeuValTyrAsnCysAspTrpAlaAlaIleGlyCysGlyAsnIleArgAlaSer 940
DB 2761 TCGACACTGGTGTACAACTGTGACTGGCGCGCATCGGGTGGGAATATATCCCGCGCTCG 2820
QY 941 PheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 956
DB 2821 TTCGGCTCGGTGAACCGCGGACGCGGACGCGGACACCTACCTCGAG 2868

RESULT 3

US-10-156-761-1

; Sequence 1, Application US/10156761

; Publication No. US20030119018A1

; GENERAL INFORMATION:

; APPLICANT: OMURA, SATOSHI
 ; APPLICANT: IKEDA, HARUO
 ; APPLICANT: ISHIKAWA, JUN
 ; APPLICANT: HORIKAWA, HIROSHI
 ; APPLICANT: SHIBA, TADAYOSHI
 ; APPLICANT: SAKAKI, YOSHIYUKI
 ; APPLICANT: HATTORI, MASAHIRA
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-262
 ; CURRENT APPLICATION NUMBER: US/10/156,761
 ; CURRENT FILING DATE: 2002-05-29
 ; PRIOR APPLICATION NUMBER: JP 2001-204089
 ; PRIOR FILING DATE: 2001-05-30
 ; PRIOR APPLICATION NUMBER: JP 2001-272697
 ; PRIOR FILING DATE: 2001-08-02
 ; NUMBER OF SEQ ID NOS: 15109
 ; SEQ ID NO 1
 ; LENGTH: 9025608
 ; TYPE: DNA
 ; ORGANISM: Streptomyces avermitilis
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (4187715)
 ; OTHER INFORMATION: a, t, c, g, other or unknown
 ; US-10-156-761-1

Alignment Scores:

Pred. No.: 2,76e-166 Length: 9025608
 Score: 2487.00 Matches: 490
 Percent Similarity: 64.77% Conservative: 135
 Best Local Similarity: 50.78% Mismatches: 248
 Query Match: 48.43% Indels: 93
 DB: 15 Gaps: 18

US-09-917-376-1 (1-957) x US-10-156-761-1 (1-9025608)

Qy 9 LeuThrMetArgSerArgLeuValSerLeuAlaAlaThrAlaSerPheAlaVal 28
 Db 2277388 CTCACCGTGGAGAACCCGCACTTC-----ACGGCGGTGCTCGCGCTG 2277432
 Qy 29 AlaAlaAlaLeu-----GlyValLeuProIleAlaIleThrAlaSerProAlaHisAla 46
 Db 2277433 GCGCGCGGTCTGCCCGCGGACCCCGCGGCACTGGCGCGGAGCGGCCCCACGCGGACG 2277492
 Qy 47 AlaThrThrGlnProTyrThrTrpSerAsnValAlaIleGlyGlyPheValAsp 66
 Db 2277493 ATGCCCGCGGACACGTACAGCTGGAGAAACGCCCGCGTCCGACGCGCGGCTTCGTCCCC 2277552
 Qy 67 GlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGly 86
 Db 2277553 GGCATCGTCTTCAACCGCTCCGAGAGAACTCGCTACGCCCGGACCGACATCGGCGGC 2277612
 Qy 87 MetTyrArgTrpAspAlaAlaAsnGlyArgTTrpIleProLeuLeuAspTrpValGlyTTrp 106
 Db 2277613 GCCTACCGCTGGCGCGAGTCTCGAAGACCTGCGACGCGGCTGCTCGACTCGGTGGCTGG 2277672
 Qy 107 AsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLys 126
 Db 2277673 AGCGATGGGGGACACGCGGTGTGAGCGCTCGGAGCTCGCTCGGCTCGGCTCGGCTCGGCT 2277732
 Qy 127 ValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu 146
 Db 2277733 GTGTACCGCGCGCTCGGACGTACAGACGTGGACCGCGGACCGGCGGACCGGTGGCTGCTC 2277792
 Qy 147 ArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGly 166
 Db 2277793 AGGTCCGGGACCGGGCGGAGCTGGCAGAACCGACCTGCGCTTCAAGCTGGGCGGG 2277852
 Qy 167 AsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIle 186
 Db 2277853 AACATCGGGCGGGGATGGGCGGCGCTCGCGTCCGACCGGACCGAAGAACGCGTG 2277912
 Qy 187 LeuTyrPheGlyAlaProSerGlyLysGlyLeuTyrArgSerThrAspSerGlyAlaThr 206

Db 2277913 CTGTATCTCGCGCGCCCGCAGCGGCAAGGGGCTGTGGCGGTCCAGCGACTCGGGGGCTCC 2277972
 Qy 207 TrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThr 226
 Db 2277973 TGGTCGAGGTCACCGACTTCCGACACTCGGCACCTAGTGTGACGACGCGACCGACAGC 2278032
 Qy 227 ThrGlyTyrGlnSerAspIleGlnGlyValValTTrpValAlaPheAspLysSerSer 246
 Db 2278033 AGCGGTACGCGTCCGACCAACAGGGCATCGTGTGGTCACTTCACGAGTCGACGGG 2278092
 Qy 247 SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProVal 266
 Db 2278093 TCGCGGGGAGCTCCACGCGGACGCTGACGTGCGGGGTGCGCGACAGGACAACTCCGTC 2278152
 Qy 267 PheTrpSerArgAspGlyGlyValAlaThrTrpGlnAlaValProGlyAlaPheThrGlyPhe 286
 Db 2278153 TATCGCTCCAGGACGCGGGCGGACCTGGTCCCGCTGGCGCGCAGCCACCGGCCAT 2278212
 Qy 287 IleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsn 306
 Db 2278213 CTCGCCCAAGGGGCTGCTGACGCGCGGAAACGGCTGTCTGTACTCTCGGTACAGCGAC 2278272
 Qy 307 ThrGlyGlyProTyrAspGlySerSerGlyAspValTTrpLysPheSerValThrSerGly 326
 Db 2278273 AAGGCGGACCGGTACGACGCGGCAAGGACAGCTGTGGCGGTACACGACGAGACCGGG 2278332
 Qy 327 ThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyr 346
 Db 2278333 ACTGGACGAAACATCAGCCCGTTCGCGGAGCGGACAC-----TACTACGGCTTC 2278383
 Qy 347 SerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSer 366
 Db 2278384 AGCGGCTGACCGTGGACCGGACGATCCGGGACGCTGATGGCGGCTGCGTACAGCTCC 2278443
 Qy 367 TrpTrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIle 386
 Db 2278444 TGGTGGCGGACACGCGACTCTTCGCTCCACGACGCGCGGACCTTGGACGAGGCG 2278503
 Qy 387 TrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGlu 406
 Db 2278504 TGGGACTACACTCGTATCCGAGCGCTCGAACCGCTTCACCATGGATGTCTCGTCTCG 2278563
 Qy 407 ProTrpLeuThrPheGlyValGlnProAsnProProValProSerProLysLeuGlyTTrp 426
 Db 2278564 CCCTGGCTCACCTGGGAGCGAACCCCGCACCGCCCGAGACAGCCCGAACTCGGCTGG 2278623
 Qy 427 MetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGly 446
 Db 2278624 ATGACCGAGTCCCTGGAGATCGACCCGTTCCGACTCCGCGCGCATGATGTACGGAAGGCG 2278683
 Qy 447 AlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIle 466
 Db 2278684 GCGACGCTTACGCGACGCGACCACTGACGAACCTGGGACGAGCGAAGCCAGTTCACCATC 2278743
 Qy 467 AlaProMetValLysGlyLeuGluThrAlaValAsnAspLeuIleSerProProSer 486
 Db 2278744 AAGCCGATGGCGGGCGCTGGAGGAGACGCGCTCAACGACCTCGCTCGCTCCCTCC 2278803
 Qy 487 ---GlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspVal 505
 Db 2278804 GCGCGCGCCAGCTGTTTCAGCGCGCTCGGTGCATCGCGGCTCCGCGACAGGACCTC 2278863
 Qy 506 ThrAlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSerValAsp 525
 Db 2278864 ACCACGTCGCTCGCTGATGTACACGTCGCGGAACCTTCAACGAGCAGCAGCTCGAC 2278923
 Qy 526 TyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSerSerGln 545
 Db 2278924 TACGCCGAGACCGCGCGGCGGCTGGTGGCGGTTCGCAATCTCGAC-----TCGGGT 2278977
 Qy 546 ProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySer 565

Db 2278978 CCG-----CATGTGGCTTCTCGACGGACACACGGCGCCCACTGGTTCCGGGGGGCG 2279028
Qy 566 GluProGlyValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPhe 585
Db 2279029 GACCTTCGGGGGTAGCGGGGTGGACGGTTCGGCGCGCTCCGACGGCAGTGGCTTC 2279088
Qy 586 ValTrpAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAenSer 605
Db 2279089 GTGTGGAGCCGGCGGCGCACCGGG-----GTGCAGTACACGACCGGGTTCGGCACCCTCG 2279142
Qy 606 TrpAlaAlaSerGlnGlyValProAlaAlaAsnAlaGlnIleArgSerAspArgValAsnPro 625
Db 2279143 TGTCTCGCGCTTCGGCGGCTTCCTCCCGCGGGCGATCTCGATCCGACCGGGTCGACCCG 2279202
Qy 626 LysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThr 645
Db 2279203 AGACCTTCTACGGCTTCAAGTCCGGCAGGTTCCTAGTCTCGACGGCGGGGGCAGC 2279262
Qy 646 PheGlnProValAlaAla---GlyLeuProSerSerGlyAlaValGlyValMetPheHis 664
Db 2279263 TTCACGGCGTCCGCGCCACCGGCTCCGAGCGGCGACAGC-----GTGCGCTTCAAG 2279316
Qy 665 AlaValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSer----- 679
Db 2279317 GCGCTCCCGGACGACGAGGCGGACATCTGGCTGGCGGGCGCGGACGCGCGGTAC 2279376
Qy 680 GlyLeuTyrHisSerThrAsnGlyLysSerTrpSerAlaIleThrGlyValSerSer 699
Db 2279377 GGGCTGTGGCACTCGACGAGCGGCGGCGGCTTCCACCAAGTCCGCCACCGTCGACCAG 2279436
Qy 700 AlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheVal 719
Db 2279437 GCGGACACCATCGCTTTCGCAAGGCGGCGACCGGCGCTCGTACACGACGCTCTACACC 2279496
Qy 720 ValGlyThrIleGlyValThrGlyAlaTyrArgSerAspAspCysGlyThrTrp 739
Db 2279497 AGCGCAAGATCGCGGTGTGCGCGGATCTTCGGTTCGACCGCAAGGCGCGAGCTGG 2279556
Qy 740 ValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyAlaIleThrGlyAsp 759
Db 2279557 ACCCGGTCAACGACGATGCCACGATGGGTGGAGCGGCGGCGGATCACCGGTGAC 2279616
Qy 760 HisAlaAsnLeuArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAsp 779
Db 2279617 CCAGGGTCTACGGCGCGGTATGTGTACGACGAGCGGCGCGGATCGTACGGCGAC 2279676
Qy 780 IleGlyGlyAlaProSerGly-----SerProSerProSerValSerProSerAla 796
Db 2279677 ACCCGGGTCTTTCGACGCGGCGGTATCGGAGCGGCGGCGGCGGCGGCTGACCGGTG 2279736
Qy 797 SerProSerLeuSer-ProSerProSerProSerProSerProSerProSerProSe 815
Db 2279737 ACATACAGGATACGACACGATGGTGGCGGCTTCCAGGCGGATGACAGTTCGCCAAC 2279796
Qy 815 rProSerSerProSerProSerSer-----Pr 824
Db 2279797 ACCGGGTGACCGCTGGAGCGGTGTCTGGGTGTGTCTGGTTCGGTACGGCGAGAG 2279856
Qy 824 oSerProSer----- 827
Db 2279857 GTCAACAGCTGTGGAAAGCCTCGTACGCGACGCGGCTTCGGGGGTGACGCGACGGAAC 2279916
Qy 828 -----ProSerPro----- 830
Db 2279917 CTGGCTTGGAAACGGAGGTGGCGGCGGCTCTCGGTGAGCTTCGGGTTCACGGGAAGC 2279976
Qy 831 -----SerProSerProSerProSerProSerProSerAla-----Se 843
Db 2279977 TGTTCGGGATCCAAATGGCAGACCGACCGCTTCAAATCGCGGTGAGAACTGACACGGTC 2280036
Qy 843 rPro-----SerProSerSerProSerProSerProSerProSerProSerProSe 861
Db 2280037 GCCTGTATGCGAAGCCGCA-CGCTCGCTGATCCGAGGCGGCGGCGGCGGCGGCGGCGGCG 2280095

Qy 861 rProThrProSerSerProValSerGlyGlyValLysValGlnTyrLysAsnAsnAs 881
Db 2280096 CCCTCCGCTCTCTCCACGGCGGTAGACCGTGGCGCTG-----GGCAA 2280137
Qy 881 pSer-AlaProGlyAspAsnGlnIleLysProGlyLeuGlnValValAsnThrGlySerS 901
Db 2280138 CTCCCGTGTCTGGTGGCAGGCTTCGGGCGATCCGGGCTCCGGTGGTGGCGGCGAGTA 2280197
Qy 901 erSerValAsp 904
Db 2280198 CCGGCGACGAC 2280208
RESULT 4
US-10-156-761-1845
; Sequence 1845, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1845
; LENGTH: 2646
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2646)
US-10-156-761-1845

Alignment Scores:
Pred. No.: 2,06e-169 Length: 2646
Score: 2478.00 Matches: 473
Percent Similarity: 64.07% Conservative: 128
Best Local Similarity: 50.43% Mismatches: 227
Query Match: 48.26% Indels: 110
DB: 15 Gaps: 14

US-09-917-376-1 (1-957) x US-10-156-761-1845 (1-2646)

Qy 23 ThrAlaSerPheAlaValAlaAlaLeu-----GlyValLeuProIleAlaIleThr 40
Db 22 ACGGCGGTGTCTGGCGGCTGTGGCGGCGGTCTGGCGGCGGCGGCGGCGGCGGCGGCG 81
Qy 41 AlaSerProAlaHisAlaAlaThrThrGlnProTyrThrTrpSerAsnValAlaIleGly 60
Db 82 AGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 141
Qy 61 GlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrVal 80
Db 142 GCGCGGCGGCTTGGTCCCGGCGGATCGTCTTCAACCGCTCGGAGAACCTCGCTACGCC 201
Qy 81 ArgThrAspIleGlyGlyMetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeu 100
Db 202 CGCACCGACATCGGGCGGCGCTACCGTGGCGGCGGAGTCTCTCGAGACCTGGACCGCTG 261
Qy 101 LeuAspTrpValGlyTrpAsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAsp 120
Db 262 CTCGACTCGGTGGGTGGAGCGGACTGGGGGCGACAGGGGTGTCTGAGGCTCGGCTCCGAC 321

QY 474 GluGluThrAlaValAsnAspLeuIleSerProSerGlyAlaProLeuIleSerAla 493
 Db 1339 GAGGAATTCCTCGTCAGGACCTGGGCTCTGCACCGCGGAGAGCTATTATGGCGCA 1398
 QY 494 LeuGlyAspLeuGlyGlyPheThrHisAla-----AspValThrAlaValProSer 510
 Db 1399 GTCCGAGACGACACCGCTTCACCTTTGCCAGGAGAACACCTCGGGACATCGCCGAC 1458
 QY 511 ThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyrAlaGluLeuAsn 530
 Db 1459 ACGTCTGGGCAACGCCACATCGGCCACCTCGACGAGCTCGACTACGCCGGGAACCTG 1518
 QY 531 ProSerIleValAlaGlySerPheAspProSerSerGlnProAsnAspArgHis 550
 Db 1519 GTCAAGAGGCTGTCGCGCTCGCAACACCGCGGACGCAACAG----- 1563
 QY 551 ValAlaPheSerThrAspGlyGlyAsnTyrPheGlnGlySerGluProGlyGlyVal 570
 Db 1564 GTGGCCATCTCGTCCAGCGCGCGCGAGCTGGAGCATCGACTACCGCGCGACAGTCC 1623
 QY 571 ThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTyrAlaProGly 590
 Db 1624 ATGAAGCGCGGACGCTGCTGCTATTCGCGCAGCGGACGACGATCTCTGTCGACGCC 1683
 QY 591 AspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGln 610
 Db 1684 TCGTCCGCG-----GTGCAGCGCTCGCAGTTCAGGCGAGCTTTGCTCGGTCTCG 1734
 QY 611 GlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProIysThrPheTyrAla 630
 Db 1735 AGCTCTCGCGCGCGCGCTGCTCGCTCGCAAGAACAGACCAACGCGCTCTTCTACGCC 1794
 QY 631 LeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProValAla 650
 Db 1795 GGTCTCGGATCGACCTTTTACGTGCAAGAGCACCGGCGAGCTTC-----ACG 1845
 QY 651 AlaGlyLeuProSerSerGlyAlaValGlyValMet-----PheHisAlaValProGly 668
 Db 1846 CGCGGG---CCCAAGTGGGCGGCGGAGGACGATCGGGATATCGTGTCTACCCGACC 1902
 QY 669 LysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSerThrAsnGlyGly 688
 Db 1903 ACCGCGGCGACGTTGATGTCTCGACGCGCTCGGATATTCGCTCCACAGACTCGGGC 1962
 QY 689 SerSerTrpSerAlaIle---ThrGlyValSerSerAlaValAsnValGlyPheGlyLys 707
 Db 1963 ACACCTTTGGCCAACTCTCCACCGCCCTGACCAACACCTTACAGATCGCCCTGGGTG 2022
 QY 708 SerAlaProGlySerSerTyrProAlaValPheValValGlyThrIleGlyValThr 727
 Db 2023 GGCTCA---GGCTCGAATGG---AACCTGTATGCTTCCGACCC-----GGCCCGTCA 2070
 QY 728 GlyAla-----TyrArgSerAspAspCysGlyThrThrTrpValLeuIleAsnAspAsp 745
 Db 2071 GGGGCTCGCTCTACGCCAGTGGAGACAGCGCGGCTCTCTGGAGGACATCCAGGGCTCC 2130
 QY 746 GlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaValLeuArg 765
 Db 2131 CAGGGCTCGGCTCCATCGACAGCACCAAGGTGCGCGGCGGAGCGGCGGCGGCA 2190
 QY 766 ValTyrIleGlyThrAsnGlyArgGlyIleValTyr-----GlyAspIleGlyGlyAla 783
 Db 2191 GTCTAGTGGGACCCACCGCGCGGCGCTTTTACGCTCAGGGAACCGTCCGCGCGCGC 2250
 QY 784 ProSerGlyProSerProSerProSerValSerProSerAlaSerProSerLeuSerProSer 803
 Db 2251 ACGGCGGCGACT-----TCTCTCGTCGACCAACAGCAGC 2283
 QY 804 ProSerProSerSerProSerProSerProSerProSerProSerProSerSerSer 823
 Db 2284 AGCAGCAGTACCTCTTCCGCCAGCTCGAGCACCAACGCTGAGGTGCGAGCGTTGATCCAC 2343
 QY 824 ProSerProSerProSerProSerProSerProSerProSerProSerProSerAlaSer 843

Db 2344 ACCCGGGCTTCGACGGTGACTTCGTGAGGAGCACCTCGGCGCGGCTCCACGGGGTCA 2403
 QY 843 rProSerPro 846
 Db 2404 GGGGTGCGCG 2413
 RESULT 6
 US-10-026-994-1
 ; Sequence 1, Application US/10026994
 ; Publication No. US20030113732A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Dunn-Coleman, Nigel
 ; APPLICANT: Goedegebuur, Frits
 ; APPLICANT: Ward, Michael
 ; APPLICANT: Yao, Jian
 ; TITLE OF INVENTION: EVI Endoglucanase and Nucleic Acids
 ; TITLE OF INVENTION: Encoding the Same
 ; FILE REFERENCE: GC698
 ; CURRENT APPLICATION NUMBER: US/10/026,994
 ; CURRENT FILING DATE: 2002-04-30
 ; NUMBER OF SEQ ID NOS: 4
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 1
 ; LENGTH: 2710
 ; TYPE: DNA
 ; ORGANISM: Trichoderma reesei
 US-10-026-994-1
 Alignment Scores:
 Pred. No.: 2,31e-108 Length: 2710
 Score: 1631.50 Matches: 353
 Percent Similarity: 57.23% Conservative: 130
 Best Local Similarity: 41.82% Mismatches: 297
 Query Match: 31.77% Indels: 64
 DB: 15 Gaps: 23
 US-09-917-376-1 (1-957) x US-10-026-994-1 (1-2710)
 QY 25 SerPheAlaValAlaAlaAlaLeuGlyValLeuProIleAlaIleThrAlaSerProAla 44
 Db 50 TCTCGAGTCTTGTGCTTGTCTGCGGGCGCTC-----ATCCCTGCC 91
 QY 45 HisAlaAlaThrThrGlnProTyrThrTrpSerAsnValAlaIle---GlyGlyGlyGly 63
 Db 92 CATCTGCC-----TTTTCATGGAAGAACGTCACCTCGGCGCGCGCGCGC 139
 QY 64 PheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAsp 83
 Db 140 TTCTGTCCTCCGCGCATCTTCATCCCAAGACAAAGGCGTAGCATATGCACGACAGAT 199
 QY 84 IleGlyGlyMetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrp 103
 Db 200 ATTGGCGGGGTGTACCGGCTCAAC---GCCGACGACTCATGACCGCGCTCAGGATGG 256
 QY 104 Val-----GlyTrpAsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAla 119
 Db 257 ATTCTGATTAATGCGGCTGGCAACATGG-----GGCATCGACCTGTGGCTT 307
 QY 120 AspProIleAsnThrAsnLysValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAsp 139
 Db 308 GATCCGAGGACGATCAAAAGTGTATGCCGAGTCGTCATGATACGACAGCTGGAT 367
 QY 140 ProAsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThrPro 159
 Db 368 CCGAGTAATGGAGCATCATTCGCTCGTCAGACCGCGGCGCAACGCTGCTTCCACCAAC 427
 QY 160 LeuProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaVal 179
 Db 428 TTGCGCTTCAAAAGTCGGGGGTAAATCCAGGACGCGGAGCGGAGCGTCTGGCTGTC 487
 QY 180 AspProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArg 199

488 GATCGGCGCACTCCAAACATCATCTTCTTGGTGTCTCGCTCAGGAAACGGCCTCTGGAAG 547
200 SerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnProAspValGlyThrTyr 219
548 TCTACGACGGCGGCGTGTACCTTTTCAAGGTCTCGTCTCAGGCAACTGGGACGTAC 607
220 IleAlaAsnProThrAspThrGlyTyrGlnSerAspIleGlnGlyValValTrpVal 239
608 ATCCAGACCCGAGTATCCCAACGGCTACAACAGCGCAACAGCAAGCAAGTCTATGTGGT 667
240 AlaPheAspLysSerSerSerLeuGlyGlnAlaSerLysThrIlePheValGlyVal 259
668 AGCTTTCAGTCAACAGCAGCAGCAGCAGCGGGGAGCCACCTCTCGTATCTTTGTGGCAG 727
260 AlaAspPro---AsnAsnProValPheThrArgAspGlyAlaThrTrpGlnAla 278
728 GCTGATAACATCACTGCTTTCAGTCTATGTGAGCAGCAAGATGCGGCTCCAGTGGAGTCT 787
279 ValProGlyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsnHis 298
788 GTACCGGGGCGCAGCGGGAATACTTTCTCTCAAGCGCGAAACTGCGAGCCAGCAGAGAAG 847
299 ValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspVal 318
848 GCCTTGTATCTCACTTATTCGATGTCACAGGCGCGTATGATGGACACTTGGCTCAGTG 907
319 TrpLysPheSerValThrSerGlyThrTrpThrArgIleSerProValProSerThrAsp 338
908 TGGAGGTACGACATTCGAGGGGAACTTGGAAAGACATCAACCTCTCTCTGGATCAGAT 967
339 ThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThr 358
968 CTA-----TACTTTGGCTTTGGCGGCTTGGCTCGATTTGCAAAAGCCAGCAACC 1018
359 IleMetValAlaThrGlnIleSerTrpProAspThrIleIlePheArgSerThrAsp 378
1019 CTTGTGTGTGCTTCTTGAACCTCTGGTGCCAGATGCTCAGCTGTTTGGTGCAGCCGAC 1078
379 GlyGlyAlaThrTrpThrArgIleTrpAspThrTrpSerTyrProAsnArgSerLeuArg 398
1079 TCTGGGACACATGAGCCGATCTGGGCGTGGCGAGCTATCCGACTGAGACCTATTAC 1138
399 TyrValLeuAspIleSerAlaGluProTrpLeuThrPheGly---ValGlnProAsnPro 417
1139 TACAGCATCTCAACTCCCAAGCAGCGTGGATCAAGAACAACTTTATCGATGTGACGAGC 1198
418 ProValProSer-----ProLysLeuGlyTrpMetAspGluAlaMetAlaIle 433
1199 GAGTCACCGTCCGATGCTCTCAAGCGCCTCGGCTGGATGATGAGTCTCTCGAGATT 1258
434 AspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsn 453
1259 GACCCACGACGACGACACCTAGCTCTACGACCGGATGAGATCTTTGGCGGCCAC 1318
454 AspLeuThrLysTrpAspSerGlyGlnIleHisIleAlaProMetValLysGlyLeu 473
1319 GATCTCACCAACTGGGACACGCGCAACATGTGTCAATCAATCACTGGCAGCGGCATC 1378
474 GluGluThrAlaValAsnAspLeuIleSerProProSerGlyAlaProLeuIleSerAla 493
1379 GAGGAATTCCTCGCTCAGGACCTGGCTCTGCACCCGGGGAAGCGAGCTATTGGCGGCA 1438
494 LeuGlyAspLeuGlyGlyPheThrHisAla-----AspValThrAlaValProSer 510
1439 GTCCGAGACGACCAACGGCTTTCATTTGGCCAGCAGAAACGACCTCGGACATCGCCGAC 1498
511 ThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyrAlaGluLeuAsn 530
1499 AGCTTCTGGGCAACGCCACATGGGCCACTCTCGACGAGGTCTGACTACGCGGGAACCTCG 1558
531 ProSerIleIleValArgAlaGlySerPheAspProSerSerGlnProAsnAspArgHis 550
1559 GTCAAGAGGCTCTCGCTCGGCGTGGCAACACACCGCGGCGCAGCAACAG----- 1603

551 ValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGlyVal 570
1604 GTGCCATCTCGTCCGACGGCGCGCGCATCGATCGATCGCGCGCCGACGACGTC 1663
571 ThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaProGly 590
1664 ATGACCGCGCGCACGGTGGCTATTTCGCCCGCAGCGCAGCAGTCTCTGTGTCGCGCC 1723
591 AspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGln 610
1724 TCGTCCGCG-----GTGCGCGCTCGCAGTTCAGGGCAGCTTTGGCTCCGTCG 1774
611 GlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyrAla 630
1775 AGCTGCGCGCGCGCGCTCATCGCTCGCAAGAACAGACCAACAGAGCTCTTCTACGCG 1834
631 LeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProValAla 650
1835 GGTCTCCGATCGACCTTTTACGTCAAGCAAGCACCGCGCAGCAGCTTC-----ACG 1885
651 AlaGlyLeuProSerSerGlyAlaValGlyValMet-----PheHisAlaValProGly 668
1886 CGCGGG---CCAGCTGGGCGCGCAGCGCAGTCCGGGATATCGCTGCTCACCGACC 1942
669 LysGluGlyAspLeuTrpLeuAlaSerSerGlyLeuTyrHisSerThrAsnGlyGly 688
1943 ACCGCGGCGACGTGTATGTCTCAGCAGCGCTCGGCATATTCGCTCCACAGACTCGGCG 2002
689 SerSerTrpSerAlaIle---ThrGlyValSerSerAlaValAsnValGlyPheGlyLys 707
2003 ACGACCTTTGGCAAGTCTCCACCGCTGACCAACACTACAGATCGCTGGGTGTG 2062
708 SerAlaProGlySerSerTyrProAlaValPheValValGlyThrIleGlyValThr 727
2063 GGCTCA---GGCTCGAAGTGG---AAGCTGTATGCTTCGGCACC-----GGCCGTCA 2110
728 GlyAla-----TyrArgSerAspCysGlyThrTrpValLeuIleAsnAspAsp 745
2111 GGGGCTCGCTCTACGCGCAGTGGAGACAGCGCGGCTCTCTGGACGACATCCAGGGCTCC 2170
746 GlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgArg 765
2171 CAGGCTTCGCTCCATCGACAGACCAAGGTGCGCGCAGCGCAGCAGCCCGCGCAA 2230
766 ValTyrIleGlyThrAsnGlyValGlyValTyr-----GlyAspIleGlyGlyAla 783
2231 GTCTACGTGGCACCAACGGCGGGGCTCTTTTACGCTCAGGGAACCGTCCGCGCGCGC 2290
784 ProSerGlySerProSerProSerProSerProSerProSerProSerProSerProSer 803
2291 ACGGCGGGGACT-----TCCTCTCGACCAAGCAGACG 2323
804 ProSerProSerSerProSerProSerProSerProSerProSerProSerProSerSer 823
2324 AGCAGCAGTACTCTCTCCGCGCAGCTCGAGCAGCAGCTGAGGTCTGTATCCAG 2383
824 ProSerProSerProSerProSerProSerProSerProSerProSerProSerProSer 843
2384 ACCCGGCTTCGACGGTGTCTCTGTCGAGGACAGCTCGCGCGCGCTCCACGGGCTCA 2443
843 rProSerPro 846
2444 GGGGTCGCGC 2453

RESULT 7

US-10-156-761-1/c
; Sequence 1, Application US/10156761
; Publication No. US20030115018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN.


```
Db 2023 ATTACAGGTGGCTGCGACGAGACATCACCGCGCTTACCGCTCCGACGCGCGCC 2082
Qy 737 Thr-ThrTyrValLeuIleAsnAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIle 756
Db 2083 AGACATGGTGGCTGCGATCAACGACGCGCCACGACGAGTGGGGTGGATCGGGAGCGCGTC 2142
Qy 757 ThrGlyAspHisAlaAsnLeuArgValTyrIleGlyThrAsnGlyArgGlyIleVal 776
Db 2143 GTCGGTGACCGCGCATCCACGCGCGGTCTACTCTGCCACCAACGCGCGGCATCCAG 2202
Qy 777 TyrGlyAsp 779
Db 2203 TACGGGGAG 2211

RESULT 9
US-09-927-827-21
; Sequence 21, Application US/09927827
; Publication No. US20030036176A1
; GENERAL INFORMATION:
; APPLICANT: Bower, Stanley G.
; APPLICANT: Ramsauer, Thomas M.
; TITLE OF INVENTION: Directed Genetic Engineering of Xanthomonas campestris
; FILE REFERENCE: 38-10(15824)B
; CURRENT APPLICATION NUMBER: US/09/927,827
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: US 60/279,493
; PRIOR FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 69
; SEQ ID NO 21
; LENGTH: 3668
; TYPE: DNA
; ORGANISM: Xanthomonas campestris
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1001)..(2668)
US-09-927-827-21

Alignment Scores:
Pred. No.: 9,15e-75 Length: 3668
Score: 1167.50 Matches: 299
Percent Similarity: 48.55% Conservative: 137
Best Local Similarity: 33.30% Mismatches: 351
Query Match: 22.74% Indels: 112
DB: 10 Gaps: 28

US-09-917-376-1 (1-957) x US-09-927-827-21 (1-3668)
Qy 23 ThrAlaSerPheAlaValAlaAlaLeuGlyValLeuProIleAlaIleThrAlaSer 42
Db 1136 ACCACACAGGAGCATGGCTGGCTGTCTTGAGCGCTGCTG---CTGTGCTCTTTTCGACG 1192
Qy 43 Pro-----AlaHisAlaAlaThrGlnProTyrThrTrpSerAsnValAla 58
Db 1193 CCCAGCGTGGCTGGCGCGGACCGCCGACGCTCCGCGGCCCTACCAAGTGGCGAGTGTGCGCC 1252
Qy 59 IleGlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeu 78
Db 1253 ATTGGCGGTGGCGCTTTGTTCACCGGTGTGCTGTGTTCATCCCGCGCAAGTGTGCGCC 1312
Qy 79 TyrValArgThrAspIleGlyGlyMetTyrArgTrpAspAlaAlaAsnGlyArgTrpIle 98
Db 1313 TATCGCGCACCGATGTGGGTGGCGGTACCGCTGGATGGCGAGCGCGCAGCAGTGGACC 1372
Qy 99 ProLeuLeuAspTrpValGlyTyrAsnAsnTrpGlyTyrAsnGlyValValSerIleAla 118
Db 1373 GCGCTGACCGACTGGTTGGCGGTGACGACTGGAACTGTATGGGCATCGACGATCGCC 1432
Qy 119 AlaAspProIleAsnThrAsnLysValTrpAlaValGlyMetTyrThrAsnSerTrp 138
Db 1433 GTGACCCCGCGATGCGATGCGCTGTATCTGCGCGCGGCACCTATATGATGAA--- 1489
Qy 139 AspProAsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThr 158
```

```
Db 1490 CGCGCCGGGCACTGCCGCGAGTGTGGCTCTCAACCGCGCGCGCACGTTCCAGCGCTGCC 1549
Qy 159 ProLeuProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGlyLeuAla 178
Db 1550 GACCTGGCTTTAGCTGGGTGTAAACGAGCTGGGCGCGGCCAATGGCGAGCGGCTGGCG 1609
Qy 179 ValAspProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTrp 198
Db 1610 GTGACCGCGACGATGGCGCGGTGTCTGTCTGGGCTCGCGCGATGCC---GGCTGTGG 1666
Qy 199 ArgSerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThr 218
Db 1667 CGTAGCGACGATCGCGCGCGCACTGGGCGAAGTGGCGTCTTCCGAGCACCGCGCGTGG 1726
Qy 219 TyrIleAlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTrp 238
Db 1727 GCCGCTGCCCGCGCGCAATCATGTGGGCGGAGAGCGGTG---GGGATCGCTTT 1783
Qy 239 ValAlaPheAspLysSerSerSerLeuGlyGlnAlaSerLysThrIlePheValGly 258
Db 1784 GTCGTGTTTCGACGCGCAGTGGCAACACGCGTCCGCAACACCGCGCATCTACGTGGCG 1843
Qy 259 ValAlaAspProAsnAsnProValPheTrpSerArgAspGlyGlyAlaThrTrpGlnAla 278
Db 1844 GTGTCCACCGAAGACACGCGCGCTGTATGTCTCCGAAGATGCCGCGCGCATTTGGGCACCG 1903
Qy 279 ValProGlyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsnHis 298
Db 1904 GTGCGCGGCAACACGCGCGCTGCGCGCGGAGCACATGCGCGCGGAGCGATGGGCAC 1963
Qy 299 ValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspVal 318
Db 1964 ---TGGTATCTGAGCTATGGCGACCGCGCGCGCGCGCATGATGCGCGGGGAGCGCTTG 2020
Qy 319 TrpLysPheSerValThrSerGlyThrTrpThrArgIleSerProValProSerThrAsp 338
Db 2021 TGGAAATTCACGCGCGCACAGGCGCGCTGGGTGAGATCAGCCGATTCGCGAG---CCA 2077
Qy 339 ThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThr 358
Db 2078 GCCAGTGGCGAGTGGATTCGGTGGGTGGGTGGGTGGATCCGCAACATCCGAGGTG 2137
Qy 359 IleMetValAlaThrGlnIleSerTrpTrpProAspThrIleIlePheArgSerThrAsp 378
Db 2138 CTGCTGGCGCACACCTTCGCGCGTCGCGCGCGCGCGAGCTGTATCGCAGCGTGAT 2197
Qy 379 GlyGlyAlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArg 398
Db 2198 GGTGGCAAGCATGGACGCGCTG-----TTGGCGCGAT 2230
Qy 399 TyrValLeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnProPro 418
Db 2231 CGCGTGTTCGATCAACGCGCGCGCGCGTGG---ACCGCATGCGCCGCGCGCGAC----- 2281
Qy 419 ValProSerProLysLeuGlyTyrMetAspGluAlaMetAlaIleAspProPheAsnSer 438
Db 2282 -----TGGATGGGG---GGCTGGCGATCGATCCGTTCCGCGCG 2317
Qy 439 AspArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrp 458
Db 2318 AACCATGCGCTGTTCGTACCGCGCTACGCGCATCTGGCGCTCGCGCAATCTG----- 2368
Qy 459 AspSerGlyGlyGlnIleHisIleAlaPro-----MetVal 470
Db 2369 -----CAGGATTTCCGCGCGCGCGCGCGCGCGCTCGAGTGGGTTCAGGAC 2416
Qy 471 LysGlyLeuGluThAlaValAsnAspLeuIleSerProProSerGlyAlaProLeu 490
Db 2417 CGTGGGTGGAGAAACCGTGGCGCTGGACCTTCGACGCCCATGCTGCGCGCGCATCTG 2476
Qy 491 IleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAlaValProSer 510
```

Db 2477 CTACGCGCGCTCGCGATATCGACGGCTTCGGCATGACGACCTGGACCGGTG---CAG 2533
Qy 511 ThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyrAlaGluLeuAsn 530
Db 2534 TTGCAGTACGCGCGCGCCACGCGCTGACCAATGTCGGAAGCATCGATGCGCGCGCAGCGC 2593
Qy 531 ProSerIleValAlaGlySerPhe-AppProSerSerGlnProAsnAspArgHi 550
Db 2594 CCGCAGTGGGTGGCGACGGTACCGTGGCGCGCGCGCGCAACACGAAATCCGCGCGC 2653
Qy 550 sValAlaPheSerThrAspGlyGlyLeuAsnTrpPheGln---GlySerGluProGlyGlu 569
Db 2654 C---TGTAATACCGTGTGCGCGACGCAATGCTGCTTTGCGAGTGGCGCGCGCAGC 2710
Qy 569 yValThrThrGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaPr 589
Db 2711 CGGCGAGCGCGCGCGAGCATTCGCTTGGTGGCGATGCGCGCGAGTGTGTGGGACC 2770
Qy 589 oGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAla 609
Db 2771 GCGCGCGCGCGCAAT---TGGCGACGCTCGACCTTCGCGCGCGCAGTGGCAGCGTGT 2824
Qy 609 rGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTy 629
Db 2825 GAACGCGCTGCGCAACCCCGCGTGGTGTGACCGATCGGCTGGACGACGCGCTGGTA 2884
Qy 629 rAlaA---LeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGlu 647
Db 2885 TGGCGTGGATGTCGCGCGCGGAGCTGTACGAGACGACCGATGCCCGCGCGAGTTTCG 2944
Qy 647 nProValAlaAlaGlyLeu-----GGCTGTGATGCGCTGGCAGGACGGT-----GGCT 3082
Db 2945 T-----GCGACGGGTGTACAGGTGGCGCGCGCGCGGATGAGCGCACCGCGCGCA 2998
Qy 654 -----ProSerSerGlyAlaValGlyValMetPheHisAlaValProGlyLysGluGl 671
Db 2999 GCTCGCTCGCGAGCCCGTGGCGCGCAGCGTGTGTATCTGCCAGCGCCCGCGCAAG--- 3053
Qy 671 yAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSerThrAsnGlyGlySerSerTr 691
Db 3054 -----GGTGTGATGCGCTGGCAGGACGGT-----GGCT 3082
Qy 691 pSerAlaIleThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGl 711
Db 3083 GCAGTGTCTACACCGCGACGACGCTCGCTGGGATCGCGAGCGCTGGCGCGC 3142
Qy 711 ySerSerTrpAlaValPheValGlyThrIleGlyValThrGlyAlaTyrAr 731
Db 3143 CGGTGGCGCGCGCGCTGTATCTGGCGCGCGCGTGCAGCGCGTGTGATGGCGGTTCG 3202
Qy 731 qSerAspAspCysGlyThrTrpValLeuIleAsnAspGlnHisGlnTyrGlyAs 751
Db 3203 CTCGACGATGGCGCGCGCGCAGTGGCGCGCATCAACGATGACGCGCACCGCTTCGCGG 3262
Qy 751 nTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgValTyrIleGlyThrAs 771
Db 3263 C---CGTACAGCGTACCGGTGATCGCGCATTCGCGCGCGTGTACTTCGCCACCG 3319
Qy 771 nGlyArgGlyIleValTyrGlyAspIleGlyAlaProSerGlySerProSerProse 791
Db 3320 CGGCGCGCGCATTTCTACGCGCATCCGAGATGAGCAGTGCATGCTGCTCGCGA---- 3374
Qy 791 rValSerProSerAlaSerProSerLeuSerProSerProSerProSerSerPro--- 810
Db 3375 ---CGTCTGCTGTCTTACCGTGTACGCGCGCGCGCGCTGCGCGCGCGCGCGT 3430
Qy 811 -----SerProSerProSerProSerProSerProSerProSerProSerPro--- 827
Db 3431 ACACCGCGACCTGTGAGATACCGGATGACCGGT---CCACATTTCCCGCGCTGCTTG 3489
Qy 828 -----ProSerProSerProSerProSerProSerProSerProSerPro--- 836
Db 3490 CTGGGTGTGTGTCGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTGT 3549

Qy 837 -----SerProSerProSerAlaSerProSerProSerProSerSerSerPr 850
Db 3550 GCCGATGCGCGGTGCTGCGACGGCATGAGGCAATGCGGTATGGGGCTGGGCCACGCC 3609
Qy 850 oSerProSerSer-----SerProSerSerSerProSerProSerPro 864
Db 3610 GCGCGCGCAATCAGCGTGAGTTTCGATGCGAAGCAGCGCCACGCCAAGACCG 3661
RESULT 10
US-10-395-241-11
; Sequence 11, Application US/10395241
; Publication No. US20040038367A1
; GENERAL INFORMATION:
; APPLICANT: YAOI, Katsuro
; APPLICANT: MITSUISHI, Yasuishi
; TITLE OF INVENTION: NOVEL XYLOGLYCAN OLIGOSACCHARIDE-DEGRADING ENZYME, POLYNUCLEOTIDE
; FILE REFERENCE: Q73756
; CURRENT APPLICATION NUMBER: US/10/395,241
; PRIOR APPLICATION NUMBER: JP 2002-83433
; PRIOR FILING DATE: 2002-03-25
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 11
; LENGTH: 2646
; TYPE: DNA
; ORGANISM: Geotrichum sp. M128
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (120)..(2558)
US-10-395-241-11

Alignment Scores:
Pred. No.: 6.89e-72 Length: 2646
Score: 1125.50 Matches: 287
Percent Similarity: 49.82% Conservative: 131
Best Local Similarity: 34.21% Mismatches: 294
Query Match: 21.92% Indels: 127
DB: 13 Gaps: 32

US-09-917-376-1 (1-957) x US-10-395-241-11 (1-2646)

Qy 17 ValSerLeuLeuAlaAlaThrAlaSerPheAlaValAlaAlaLeuGlyValLeuPro 36
Db 114 TTAGATATGGTGGCAGTCACCTCCCTCGGAGGCGCTTACTGCGCTTCGATTCTG--- 170
Qy 37 IleAlaIleThrAlaSerProAlaHisAlaAlaThrThrGlnProTyrThrTrpSerAsn 56
Db 171 -----GGTCTGCTCGCGTGGCGCGAGGACACTACGAGTTCAAGAT 212
Qy 57 ValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGly 76
Db 213 GTGCGATCGCGCGCGCGGTACATTACCGGGATTGTGCGCACCCAAAGACCAAGGAC 272
Qy 77 IleLeuTyrValArgThrAspIleGlyMetTyrArgTrpAspAlaAlaAsnGlyArg 96
Db 273 CTGTGTACGCGCGCAGCGACATTGGCGCGCGTACCGCTGGGACGCGACGTCGAAG 332
Qy 97 TrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyrAsnGlyValValSer 116
Db 333 TGGATCCGCTCAACGACTTTATCGAGCGCGAGGACATGAACATTATGGGCGCGAGTGC 392
Qy 117 IleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaValGlyMetTyrThr--- 135
Db 393 ATCGCGTGGACCCCAACACCCCGACAGCGGTGTACTCGCGCAGGGCGCTATGTCGCG 452
Qy 136 AsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrp 155
Db 453 GACGAGTGG-----GCGCGCTTCTATGTGTCCGAAGACCGCGCGCGCTGTT 500
Qy 156 GlnIleThrProLeuProPheLeuGlyGlyAsnMetProGlyArgGlyMetGlyGlu 175

501 ACAATCTACGAGTCGCGGTTCCGATGGGCGCAACACATGGGAGCAACAATGGCGAG 560
176 ArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLys 195
561 CGCCTCGTGTCAACCGGTTCACTCGAAGAGTCTGGATGGGTACGGGTACA---GAG 617
196 GlyLeuTyrArgSerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAsp 215
618 GGTATCTGAAGAGTTCGGACCGCGCAACCTGGACAAACGTCACTCCATCCCGGAC 677
216 ValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGly 235
678 GCGTTC-----ACCAACGGTATCGGATACACGTCG----- 707
236 ValValTyrValAlaPheAspLysSerSerSerLeuGlyGlnAlaSerLysThrIle 255
708 -----GTCATTTCGACCCC-----GAACGTAAATGACCAATC 740
256 PheValGlyValAlaAspProAsnAsnProValPheTrpSerArgAspGlyGlyAlaThr 275
741 TAGCGAGCGGACTGCCCGCGAGGGC---ATGTAGCTCACGACGACGCGGTGTCTCG 797
276 TrpGlnAlaValProGlyAlaProThrGlyPheIle----- 287
798 TGGGAGCCAGTGGCGGCGACCGCTCCAGCTGCTCAACAGGACCGCGGCGGTTCGCCG 857
288 -----ProHisLysGlyValPheAspProValAsn 297
858 GACAGAAGCCCGCTGATCGCGCGCGACGCCATGAATGCTCTCACCCCC----- 911
298 HisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAsp 317
912 AACTTCTCTACGTACTTACGCGGACTACCTCGTTCATCGGCGGTCAAGTTCGGCGAA 971
318 ValTrpLysPheSerValThrSerGlyThrTrpThrArgIle----- 331
972 GTCTGGCGCGAAGACCGACCTCGCGCGCTCGGACGACATTATCCCCCGCTCGGCAAC 1031
332 ---SerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThr 350
1032 TCGTCGCTCCCGCTCCGCTACAAACACAGAGCTTCCTCGCGGCGGATTTTCGGGTCTCAGC 1091
351 IleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrpTrpProAsp 370
1092 GTCCGCGGACCAACCCCAACCTCTCGTCTATCACC---CTCGACCGGACCCCGGA 1148
371 ThrIle-----IlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIleTrp 387
1149 CCGCGCTCGACAGCATCTACCTCTCAACCGATGCGCGCGACCTGGAAGGACGTCAAC 1208
388 AspTrpThrSer-----TyrProAsnArgSerLeuArg 398
1209 CAGCTCTCGTCCCGCTCCAACTCGAAGTAACTGGGGCCACCCGACTAACCGCGCGCG 1268
399 TyrValLeuAspIleSerAlaGluProTyrLeuThrPheGlyValGlnPro----- 415
1269 TAC---AAGACCGGACGCTGTTCGTGCTCGACTTCAACACGGTCCCGAGTGGGGG 1325
416 -----AsnProValProSerPro-----LysLeuGlyTrpMetAspGluAla 430
1326 GGATACGGTGGCGCGACGCTAGCCCGGCTCACCAAGTTTGGCTGGTGATGAGCGCT 1385
431 MetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeuTyr 450
1386 GTGCTTATCGATCCGTTCAACCCCGACCTCTGATGTACGCGACGCGGGGCGACCATCTGG 1445
451 AlaThrAsnAspLeuThrLys-----TrpAspSerGlyGlyGlnIleHisIle 466
1446 GCGACCGACAGCTCTCCCGGTGTCGAGAAGACTGG----- 1481
467 AlaPro-----MetValLysGlyLeuGluThrAlaValAsnAspLeuIle 482

1482 GCGCGAGCTGGTACCTCCAGATCGACGGTATCGAGGAGAATGCGATCCTGTGCTCCGC 1541
483 SerProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHis 502
1542 TCGCCCAAGAGCGCGCGCTCTCTGTGGGCATCGGTGACATTAGCGGCATGAAGCAC 1601
503 AlaAspValThrAlaValProSerThrIlePheThrSerProValPheThrThrGlyThr 522
1602 GAGCAGCTCCCAAG---CCCAAGAAGATGTTGGTGGCGCCAGTTCTCCAACTCCGAC 1658
523 SerValAspTyrAlaGluLeuAsnProSerIleValArgAlaGlySerPheAspPro 542
1659 AGCATGACGCTGGCGCAACTTCCCAACGTGTGTCGCGCGCGATCTCTCGGGACAC 1718
543 SerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrp--- 561
1719 GAGTACGACAGCGCGTGGCGCGTGGTACGCGACTGACGGCGGAGACGCGTGGACC 1778
562 ---PheGlnGlySerGluProGlyGlyValThr-----GlyGlyThrValAla 577
1779 ATCTTCTCTACCTGCTGCGCATGAACGCGACCTACCGGCGAGCAGCAGATTGCA 1838
578 AlaSerAlaAspGlySerArgPheValTrpAlaPro-----GlyAspProGlyGlnPro 595
1839 GTCGACCGGAGCGGACGACATCGTGTGGTCGACCAAGCTTGACGAGCGGCTCGGGA 1898
596 ValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAla--- 614
1899 CCGTGTACTCGCAGCATATGGCAAGCGTGTCT-----GTTCCCGCTGCG 1946
615 -----AsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyr 629
1947 GACCTGAAGGCCAGACTGCCAATGTCTCTCGAACAAGTCCAGGATGGCAGTTCTAC 2006
630 AlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProVal 649
2007 GTTACCGATGGCGCAAGTCTCTGCTCGACCGGCGGGAAGTATGCGGCCAAG 2066
650 AlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaValProGlyLys 669
2067 GCGCGCGGACTTGTCACT-----GGCACATCGCTCATGCTCGCGTGAACCCCTGGTG 2120
670 GluGlyAspLeuTrpLeu---AlaAlaSerSerGlyLeuTyrHisSerThrAsnGlyGly 688
2121 GCGCGCGACGCTGGTGTCTCCGAGGCGGTCTCTCCACTCGACCGACGCTTGGC 2180
689 SerSerTrpSerAlaIle---ThrGlyValSerSerAlaValAsnValGlyPheGlyLys 707
2181 GCCTCGTTACGAGGTAGTACCGCCAAACGACCCCTCGTGAGCGTGGCGCGCCCAAG 2240
708 Ser-----AlaProGlySerSerTyrProAlaValPheValValGlyThr--- 722
2241 TCCAGTCCGACGCGCAAGAGCTAGCGCGCTCGCGGTCTTTCATCTGGGGCACCGAC 2300
723 ---IleGlyGlyValThrGlyValTyrArgSerAspCysGlyThrTrpValLeu 741
2301 AAGCTCGAAGCGACATCGGCTGTACCGCTCCGACGACACGCGCAGCACCTGGACGCG 2360
742 IleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAla 761
2361 GTCAATGACGAGGACCACTACTCTCGGCG---CCCAACCATGATGAGCGCCGCCAAG 2417
762 AsnLeuArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIle 780
2418 GTTACGGCGCGGTATCTAGGACGAAGCGCGCGGTATCGTGTACCGCCACCTT 2474

RESULT 11

US-10-395-241-17
; Sequence 17, Application US/10395241
; Publication No. US20040038367A1
; GENERAL INFORMATION:
; APPLICANT: YAOI, Katsuro
; APPLICANT: MITSUISHI, Yasushi

;; TITLE OF INVENTION: NOVEL XYLOGLUCAN OLIGOSACCHARIDE-DEGRADING ENZYME, POLYNUCLEOTIDE
 ;; FILE REFERENCE: Q73756
 ;; CURRENT APPLICATION NUMBER: US/10/395,241
 ;; PRIOR FILING DATE: 2003-03-25
 ;; PRIOR APPLICATION NUMBER: JP 2002-83433
 ;; PRIOR FILING DATE: 2002-03-25
 ;; NUMBER OF SEQ ID NOS: 19
 ;; SOFTWARE: PatentIn version 3.2
 ;; SEQ ID NO 17
 ;; LENGTH: 2481
 ;; TYPE: DNA
 ;; ORGANISM: Artificial Sequence
 ;; FEATURE:
 ;; OTHER INFORMATION: Xyloglucan Oligosaccharide-Degrading Enzyme with Histidine Tag
 ;; FEATURE:
 ;; NAME/KEY: CDS
 ;; LOCATION: (1)..(2481)
 ;; US-10-395-241-17

Alignment Scores:
 Pred. No.: 8,25e-72 Length: 2481
 Score: 1124.00 Matches: 298
 Percent Similarity: 48.91% Conservative: 127
 Best Local Similarity: 34.29% Mismatches: 306
 Query Match: 21.89% Indels: 138
 DB: 13 Gaps: 33

US-09-917-376-1 (1-957) x US-10-395-241-17 (1-2481)

Qy 52 TyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsn 71
 Db 13 TACGAGTTCAGAGATGTCGGATCGCGGCGGGGTACATTACCGGGATGTCGCCAC 72
 Qy 72 GluGlyAlaProGlyIleLeuValArgThrAspIleGlyMetTyrArgTrpAsp 91
 Db 73 CCAAGACCAAGACCTGCTGTATCGCGCGCACGACATTGGCGCGGTACCGCTGGAC 132
 Qy 92 AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyr 111
 Db 133 GCAGGACGTCCTCAAGTGGATCGCGCTCAACGACTTATCAGGCGCGAGGACATGAACATT 192
 Qy 112 AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaVal 131
 Db 193 ATGGGACACGATCGATCGCGCTGGACCCCAACACCCGACAGCTGTACCTCGCGCAG 252
 Qy 132 GlyMetTyrThr---AsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAsp 150
 Db 253 GGGCGCTATGTCGGCGACGAGTGG-----GCGGCGTTCATGTGTCGAAGAC 300
 Qy 151 GluGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGly 170
 Db 301 CGCGGCGAGTCGTTTCAATCTACGAGTCCCGCTTCCGATGGCGCGCAACGACATGGGA 360
 Qy 171 ArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGly 190
 Db 361 CGAACATGGCGAGCGCTCGCTGTCAACCCCTTCACTCGAACGAGGTCTGATGGGT 420
 Qy 191 AlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMet 210
 Db 421 ACGCGTACA---GAGGATCTCGAAGAGTTCGAGCGCGCCCAAGACCTCGAACAAAGTC 477
 Qy 211 ThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGln 230
 Db 478 ACCTCCATCCGCGCGCTTC-----ACCAACGGTATCGGATACACG 519
 Qy 231 SerAspIleGlnValValTrpValAlaPheAspLysSerSerSerLeuGlyGln 250
 Db 520 TCG-----GTCAATTTTCGACCCC-----GAA 540
 Qy 251 AlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrpSerArg 270
 Db 541 CGTAATGGCACCATCTACGCGGCGGAGCTGCCCCCGCAGGGC---AGTACGTACCGCAC 597

Qy 271 AspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIle----- 287
 Db 598 GACGCGGTGTCTCGTGGAGCCAGTGGCGGCGCAGCGCTCCAGCTGGCTCAACAGGACC 657
 Qy 288 -----ProHisLysGlyVal 292
 Db 658 ACGGCGCGTTCCCGGACAAGACCCCGCTCGATCGCGCGCAGCCCATGAAGTCGCT 717
 Qy 293 PheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAsp 312
 Db 718 CTCACCCCC-----AATCTCTCTACGTGACTTACGCGGACTACCTGCTCATGGGC 771
 Qy 313 GlySerSerGlyAspValTrpLysPheSerValThrSerGlyThrTrpThrArgIle--- 331
 Db 772 GTCAGTTCGGCAAGICTGGCGCGCAGAACCGCACTCGGGCGCTCGGAGACATTACT 831
 Qy 332 -----SerProValProSerThrAspThrAlaAsnAspTyrPheGly 345
 Db 832 CCCGCGTGGCAACTCGCTCGCTGCCCTGACCAACACAGACGCTTCCTCGCGCGGA 891
 Qy 346 TyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIle 365
 Db 892 TTTTGGCGTCTCAGCGTCGACGCGACCAACCCCAACCTCTCGTCTCATCACC---CTC 948
 Qy 366 SerTrpTrpProAspThrIle-----IlePheArgSerThrAspGlyGlyAlaThr 382
 Db 949 GACGCGACCCGCGACCGCCCTCGACAGCATCTACCTCTCAACGATGCGCGCGGAC 1008
 Qy 383 TrpThrArgIleTrpAspTrpThrSer-----TyrPro 393
 Db 1009 TGAAGAGCGTCACCCAGCTCTCGTCCCGTCAACCTCGAAGGTAACTGGGCGCACCCG 1068
 Qy 394 AsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrpLeuThrPheGlyVal 413
 Db 1069 ACTAACGCGCGCGGTAC---AAGACGCGCACGCGCTGTTCGTGCTCGACTTCAACAAC 1125
 Qy 414 GlnPro-----AsnProProValProSerPro-----LysLeuGly 425
 Db 1126 GGTCCCGAGTGGGGGATACGCTGCGCGCACGATCGCCCGCTCCAGGATTTGGC 1185
 Qy 426 TrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThr 445
 Db 1186 TGGTGGATGAGCGTGTCTTATCGATCCGTTCAACCCCGAGCACCTGATGTACGCGCAG 1245
 Qy 446 GlyAlaThrLeuTyrAlaThrAsnAspLeuThrLys-----TrpAspSerGly 461
 Db 1246 GGGCGGACCATCTCGCGGACCGCACGCTCTCCGCTGTCGAGAAGACTGG----- 1296
 Qy 462 GlyGlnIleHisIleAlaPro-----MetValLysGlyLeuGluGluThrAla 477
 Db 1297 -----GCGCCGAGTGGTACTCCAGATCGACGCTATCGAGAGATGCG 1341
 Qy 478 ValAsnAspLeuIleSerProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeu 497
 Db 1342 ATCTGTGCTCCGCTCGCCCAAGCGCGCGGCTCTCTGTGGGATCGGTGACATT 1401
 Qy 498 GlyGlyPheThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProVal 517
 Db 1402 AGCGGATGAAGACGACGACCTCACCAAG---CCCCAAGAAGATGTTTGGTGGCGCCAG 1458
 Qy 518 PheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAla 537
 Db 1459 TTCTCCAACTCCGACGATCGACGCTCGCGGCAACTTCCCAACGTTGCTCGCGCGC 1518
 Qy 538 GlySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGly 557
 Db 1519 GATCTCCGGACACGAGTACGACGCGGTGGCGCGGTGCTGCTAGCGACTGACGCG 1578
 Qy 558 GlyLysAsnTrp-----PheGlnLysSerGluProGlyGlyValThrThr----- 572
 Db 1579 GGAGACGCGTGGACCATCTTCCCTACCTGCTCCCTCTGCGATGAACGCGGAGCCATACCAG 1638

573 GlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaPro-----Gly 590
1639 GGCAGCAGGATTCAGTCAGCGCGGCGGAGATCGTGGTGGCCCAACAGCTTGAC 1698
591 AspProGlyGlnProValValTrpAlaValGlyPheGlyAsnSerTrpAlaAlaSerGln 610
1699 GAGCAGGCTCGGACCGTGGTACTCGCAGCATATGGCAAGCGTGTCT----- 1749
611 GlyValProAla-----AsnAlaGlnIleArgSerAspArgValAsn 624
1750 ---GTTCCGCTGGCGACCTGAAGGCCAGCTGCTCTCGGACAAAGTCCAG 1806
625 ProLysThrPheTyAlaLeuSerAsnGlyThrPheTyArgSerThrAspGlyGlyVal 644
1807 GATGGCAGCTTCTACGCTACCGATGGCGGCAAGTTCTCTCGACCGGCGGGAAG 1866
645 ThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHis 664
1867 TGTATGCGCCGCAAGGCGCGGACTTGTCTCT-----GGCAGTGGCTCATGCTGCC 1920
665 AlaValProGlyLysGluGlyAspLeuTrpLeu-----AlaAlaSerSerGlyLeuTyHis 683
1921 GTCAACCCCTGGGTGGCGCGGAGCGTCTGGTGTCTTCCGAGGCGGTCTCTCCAC 1980
684 SerThrAsnGlyGlySerSerTrpSerAlaIle-----ThrGlyValSerSerAlaValAsn 702
1981 TCACCGAGCTTGGCGCTCGTTTCAGAGGCTAGTACCCGCAACGCGACCTCGTGAGC 2040
703 ValGlyPheGlyLysSer-----AlaProGlySerSerTyProAlaValPhe 718
2041 GTGCGGCGCCCAAGTCCAAAGTCGAGCGGCAAGAGGCTAGCGCGCCCTCCGCGGTCTTC 2100
719 ValValGlyThr-----IleGlyGlyValThrGlyAlaTyArgSerAspCysGly 736
2101 ATCTGGGACCGCAGCAAGCTGAGCGCATCGGCTGTACCGCTCCGACGCAACGGC 2160
737 ThrThrTrpValLeuIleAsnAspAspGlnHisGlnTyArgSerTrpGlyGlnAlaIle 756
2161 AGCACCTGAGCGCGCTCAATGACGAGGAGCAACACTTACTCTCGGC-----CCACCATGATC 2217
757 ThrGlyAspHisAlaAsnLeuArgArgValTyIleGlyThrAsnGlyArgGlyIleVal 776
2218 GAGCGCGACCCCAAGCTCTACGCGCGCGTGTATCTAGGCAAGCGCGCGGTATCGTG 2277
777 TyrGlyAspIleGlyAlaProSerGlySerProSerProSerValSerProSerAla 796
2278 TAGCGGACCTTACCAACAAAGAGAGC---AACGAGGAGAGTCCGCGCAAGTGGCGCC 2334
797 Ser-----ProSerLeuSer 801
2335 AACGGCCAGAGGCGACGCTGCTATGTGAAAAGGAGATCTGGGTACCTGTGTGCCAC 2394
802 ProSerProSerProSerSerProSerProSerProSerProSerProSerProSe 821
2395 GCGGTTCCATGGCTGATATCGGATCCGAATTCGAGCTCCGACAGCTTGGCGCGCA 2454
821 rSerSerProSerPro 828
2455 CTGAGCACCACCACCACCA 2477

RESULT 12

US-10-395-241-13
Sequence 13, Application US/10395241
Publication No. US20040038367A1
GENERAL INFORMATION:
APPLICANT: YAOI, Katsuro
APPLICANT: MITSUISHI, Yasushi
TITLE OF INVENTION: NOVEL XYLOGLUCAN OLIGOSACCHARIDE-DEGRADING ENZYME, POLYNUCLEOTIDE
TITLE OF INVENTION: ENCODING THE SAME, AND METHOD OF PREPARING THE ENZYME
FILE REFERENCE: Q73756
CURRENT APPLICATION NUMBER: US/10/395,241
CURRENT FILING DATE: 2003-03-25
PRIOR APPLICATION NUMBER: JP 2002-83433

PRIOR FILING DATE: 2002-03-25
NUMBER OF SEQ ID NOS: 19
SOFTWARE: Patent in version 3.2
SEQ ID NO 13
LENGTH: 2367
TYPE: DNA
ORGANISM: Geotrichum sp. M128
FEATURE:
NAME/KEY: mat_peptide
LOCATION: (1)..(2367)
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(2367)
US-10-395-241-13
Alignment Scores:
Pred. No.: 2,51e-71
Score: 1117.00
Percent Similarity: 50.37%
Best Local Similarity: 34.95%
Query Match: 21.75%
DB: 13
Length: 2367
Matches: 281
Conservative: 124
Mismatch: 279
Indels: 120
Gaps: 31

US-09-917-376-1 (1-957) x US-10-395-241-13 (1-2367)

Qy 52 TyrThrTrpSerAsnValAlaIleGlyGlyGlyGlyGlyPheValAspGlyIleValPheAsn 71
Db 10 TACGAGTTCAAGATGTCGGATCGCGGCGCGGATACATTACCGGGATTGTCGGGCAC 69
Qy 72 GluGlyAlaProGlyIleLeuTyArgValArgThrAspIleGlyGlyMetTyArgTrpAsp 91
Db 70 CCAAGACCAAGGACCTGCTGTACGCGCGCAGGACATTGGCGCGCTACCGCTGGGAC 129
Qy 92 AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTy 111
Db 130 GCAGGACGTCGAAGTGGATCCCTCAACGACTTATTCGAGCGCGGAGACATGACATT 189
Qy 112 AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaVal 131
Db 190 ATGGCCACGAGTGCATCGCGTGGACCCCAACACCCGACAGGCTGTACCTCGCGCAG 249
Qy 132 GlyMetTyThr-----AsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerAsp 150
Db 250 GGGCGCTATGTCGGCGACGAGTGG-----GCGCGCTTCTATGTGTCCGAAGAC 297
Qy 151 GlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGly 170
Db 298 CGCGGCGAGTGTGTACATCTACGAGTCCGCTTCCGATGGCGCGCCACACATGGGA 357
Qy 171 ArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyPheGly 190
Db 358 CGCAACATGGCGAGCGCTCGCTGTCAACCGTTCACTCGAACGAGTCTGGATGGGT 417
Qy 191 AlaProSerGlyLysGlyLeuTyArgSerThrAspSerGlyAlaThrTrpSerGlnMet 210
Db 418 ACGGCTACA---GAGGGTATCTGGAAGATGTCGAGCGCGCCAGACCTGGCAACAGTC 474
Qy 211 ThrAsnPheProAspValGlyThrTyIleAlaAsnProThrAspThrThrGlyTyGln 230
Db 475 ACGTCCATCCCGGACGCGTTC-----ACCAAGGTATCGGATACACG 516
Qy 231 SerAspIleGlnGlyValTrpValAlaPheAspLysSerSerSerLeuGlyGln 250
Db 517 TCG-----GTCAATTTTCGACCC-----GAA 537
Qy 251 AlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrpSerArg 270
Db 538 CGTAATGCGACCATCTACGAGCGCGAGCTGCCCGCAGGGC---ATGTACGTCCAGCAC 594
Qy 271 AspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIle----- 287
Db 595 GACGCGGTGTCTCGTGGGAGCCAGTGGCGGCGCCAGCGTCCAGCTGGCTCAACAGGACC 654

QY 288 -----ProHisLysGlyVal 292
Db 655 ACGGCGCGTTCCGGCAAGAACCCGCGTGCATCGCGCGCAGCCCGCAAGTGCCT 714
QY 293 PheAspProValAenHisValLeuTyrIleAlaThrSerAenThrGlyGlyProTyrAsp 312
Db 715 CTCACCCCC-----AACTTCCTCTAGTGAATACCGCCGACTACCTGTGGTCCATGGGCG 768
QY 313 GlySerSerGlyAspValTrpLysPheSerValThrSerGlyThrTrpThrArgile--- 331
Db 769 GTCAGGTTGGCGAAGTCTGGCGCCAGAACCGCAGCTCGGGCGCTGGGACGACATTACT 828
QY 332 -----SerProValProSerThrAspThrAlaAenAspTyrPheGly 345
Db 829 CCCCGCTGGCGCAACTCGTCGCTGCCCTGTACCAACCAACAGAGCTTCCCTCGCGGGCGA 888
QY 346 TyrSerGlyLeuThrIleAspArgGlnHisProAenThrIleMetValAlaThrGlnIle 365
Db 889 TTTTGGCGGTCTCAGCGTGCAGCGGACCAACCCCAACCGTCTCGTGCATCACC---CTC 945
QY 366 SerTrpTrpProAspThrIle-----IlePheArgSerThrAspGlyGlyAlaThr 382
Db 946 GACCGCACCCGGACCGCGCTCGACAGCATCTACTCTCAACCGATCGCGCGCGAC 1005
QY 383 TrpThrArgIleTrpAspTrpThrSer-----TyrPro 393
Db 1006 TGAAGGACGTCACCCAGCTCTGTCCTCCGTCACACCTCGAAGCTAACTGGGGCCACCG 1065
QY 394 AsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrpLeuThrPheGlyVal 413
Db 1066 ACTAACCGCGCGGTAC---AAGGACGCGACCGCTGTTCGTGGCTCGACTTCAACAAC 1122
QY 414 GlnPro-----AsnProProValProSerPro-----LysLeuGly 425
Db 1123 GGTCCCCAGTGGGGGGGATACGGTGGCGCGCGAGTACGCGCGCTCTCAACAGTTTGGC 1182
QY 426 TrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThr 445
Db 1183 TGGTGATGAGCGTGTGCTTATCGATTCGTTTCAACCCCGACGACCTGTGTAGTACGGCAG 1242
QY 446 GlyAlaThrLeuTyrAlaThrAsnAspLeuThrLys-----TrpAspSerGly 461
Db 1243 GGGCGGACCATCTGGGCGACCGACACGCTCTCCGTTGTCGAGAAGGACTGG----- 1293
QY 462 GlyGlnIleHisIleAlaPro-----MetValLysGlyLeuGluThrAla 477
Db 1294 -----GGCGCGAGCTGTACTCTCAGATCCAGCGTATCGAGGAGAAATGCG 1338
QY 478 ValAsnAspLeuIleSerProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeu 497
Db 1339 ATCTGTGCTCGCTCGCCCAAGAGCGCGCGGCTCTCTGCGGCATCGGTGACATT 1398
QY 498 GlyGlyPheThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProVal 517
Db 1399 AGCGCATGAAGCAGCAGCAGCTCACCAAG---CCCCAGAAGATGTTTGGTGGCGCCCGAC 1455
QY 518 PheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAla 537
Db 1456 TTTCTCAACCTCGACAGCATCGAGCTGGCGGCAACTTCCCAACGTTGTGCTCGCGCGC 1515
QY 538 GlySerPheAspProSerGlnProAsnAspArgHisValAlaPheSerThrAspGly 557
Db 1516 GGATCTCGGACAGATACGACGCGGTGGCGCGGTGCTGACGCGCTGACGCGC 1575
QY 558 GlyLysAsnTrp-----PheGlnGlySerGluProGlyGlyValThr----- 572
Db 1576 GGAGACGCTGGACCATCTTCCCTACTCTGCTCTCGATGAACGCGACCTACCAG 1635
QY 573 GlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaPro-----Gly 590
Db 1636 GGCAGCAGCATGTGCTGCGAGCGGCGGCGGACGATGTGTGTGTCGACCAAGCTTGAC 1695
QY 591 AspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGln 610

Db 1696 GAGCAGGCTCTGGGACCGTGGTACTCGCACGACTATGGCAAGACGTGGTCT----- 1746
QY 611 GlyValProAla-----AenAlaGlnIleArgSerAspArgValAsn 624
Db 1747 ---GTTCCCGCTGGCGACCTGAAGCCCGAGACTGCAATGTCTTCGACCAAGGTCAG 1803
QY 625 ProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyVal 644
Db 1804 GATGSCAGTTTACTACGTACCGATGGCGCAAGTCTTCTGCTCGACCGACGCGGGAAG 1863
QY 645 ThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHis 664
Db 1864 TCGTATGCGCGCAAGGGCGCGGACTTGTCACT-----GGCACATCGCTCATGCTGCC 1917
QY 665 AlaValProGlyLysGluGlyAspLeuTrpLeu---AlaAlaSerSerGlyLeuTyrHis 683
Db 1918 GTGAACCCCTGGTGGCGCGGACGCTCTGGGTGCTGTTCGCCAGGCGGTCTCTTCCAC 1977
QY 684 SerThrAsnGlyGlySerSerTrpSerAlaIle---ThrGlyValSerSerAlaValAsn 702
Db 1978 TCGACCGACTTGGCGCTCTGTTACAGAGGTAGGTACGCCCAACGCGACCTCGTGAGC 2037
QY 703 ValGlyPheGlyLysSer-----AlaProGlySerSerTyrProAlaValPhe 718
Db 2038 GTCGCGCGCCCAAGTCCAAGTCGACGCAAGAGGCTAGCGCGCTCCCGGCTTTC 2097
QY 719 ValValGlyThr-----IleGlyValThrGlyAlaTyrArgSerAspCysGly 736
Db 2098 ATCTGGGGCACCGACCAAGCTGGAAAGCAGCATCGGCTGTACCGCTCCGACGACACGCG 2157
QY 737 ThrThrTrpValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIle 756
Db 2158 AGCACTTGGACGCGGTCAATGACGAGGACAACTACTCGGCG---CCCACCATGATC 2214
QY 757 ThrGlyAspHisAlaAsnLeuArgValTyrIleGlyThrAsnGlyArgGlyIleVal 776
Db 2215 GAGCGCGACCCCAAGTCTACGGGCGCTGTATCTAGGACGACGCGCGGTATCGTG 2274
QY 777 TyrGlyAspIle 780
Db 2275 TACGCGCGACCT 2286

RESULT 13
US-10-420-191-1
; Sequence 1, Application US/10420191
; Publication No. US20040067569A1
; GENERAL INFORMATION:
; APPLICANT: No. US20040067569A1ozymes Biotech, Inc.
; APPLICANT: Rey, Michael W.
; APPLICANT: Zaretsky, Elizabeth J.
; APPLICANT: Haas, Jeffrey A.
; TITLE OF INVENTION: Polypeptides Having Xyloglucanase Activity And Nucleic Acids
; FILE REFERENCE: 10210-200-US
; CURRENT APPLICATION NUMBER: US/10/420,191
; PRIOR FILING DATE: 2003-04-18
; PRIOR APPLICATION NUMBER: US 60/373,987
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 5698
; TYPE: DNA
; ORGANISM: Tricoderma reesei
US-10-420-191-1

Alignment Scores:
Pred. No.: 3,71e-58 Length: 5698
Score: 940.00 Matches: 374
Percent Similarity: 30.11% Conservative: 136
Best Local Similarity: 22.08% Mismatches: 315
Query Match: 18.31% Indels: 871

Qy 711 GlySerSerTyrProAlaValPheValValGlyThrIleGlyGlyValThrGlyAla--- 729
Db 4413 AGTCCGAGCTTGACTTGGACATACGATCGCAC---GGCCCGTTCAGGGGCTCGC 4466
Qy 730 ---TyrArgSerAspAspCysGlyThrThrTrpVal----- 740
Db 4467 CTCTACCCAGTGGAGACACGGCGCTCTCGAAGCCGTGGCGGCGAGTCCCGGAGCG 4526
Qy 741 -----LeuIleAsnAspAspGlnHisGlnTyrGly--- 750
Db 4527 GAGATCGGTCACCTCTGTCTGCGCGCGGAGGACCTCGGACATCCAGGCTCCAGGGCTTC 4586
Qy 750 ----- 750
Db 4587 GGCTCCATCAGACACCAAGGTGCGCGGCGAGCGGCTGTAGTTCGCGAGGTCGCCGAG 4646
Qy 751 -----AsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgVal 766
Db 4647 CCGAGGTAGTGTCTGTGGTTCAGCGCGCGTTCGCGGACACCGCGCGG-----CAAGTC 4700
Qy 767 TyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAsp----- 779
Db 4701 TAGTGGGACCAACGCGCGGGGCTCTTTACGCTCAGGCGTCTGTGGCGCGGCTTCAG 4760
Qy 780 -----IleGlyGlyAlaPro 784
Db 4761 ATGCACCGCTGGTTCGCGGCGCGCGACGAAATGCGAGTCCGAAACGCTCGCGCGCGC--- 4817
Qy 785 SerGlySerProSerProSerValSerProSerAlaSerProSerLeuSerProSerPro 804
Db 4818 ACGGCGGGGACTTCTCTGTCGACCAAGCAGCAGCAGCAGTACTTGGCAGCGCGCGCG 4877
Qy 805 SerPro-----SerSerSerProSerProSerProSerPro 816
Db 4878 TGCCGCGCTGAAGGACGACTGTTCTGCTCTGCTGCTGCTCTCTTCGCGCAGTCTG 4937
Qy 817 SerSerSerProSerSerProSerProSerProSerPro----- 828
Db 4938 AGCACACGCTGAGTTCGAGCTGTGTTATCCAGACCGCGGCTTGAGAGAGCGGCTCGAGC 4997
Qy 829 -----SerProSerProSer----- 833
Db 4998 TCGTGTGCGACTCCAGCTCGCAACATAGTGTGCGGCGCGGACGACGCGTGTCTGTCG 5057
Qy 834 ---ProSerArgSerProSerProSerAlaSerProSerProSerProSerProSerPr 852
Db 5058 AGGACGAGCTGGCGCGCGCTCCACCGGGTCCAGGGTCCAGGGTCCGCGCTGCCACTGAAGCAGC 5117
Qy 852 oSerSerSerProSerSerProSerProSerProSerProSer 865
Db 5118 TCCTGTGCGAGCGCGCGCGGCGGAGGTGCCCGCCAGTCCCGCAGC 5157

RESULT 14

US-09-917-384-2
; Sequence 2, Application US/09917384
; Publication No. US20030096342A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: AUNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOXYLICUS
; FILE REFERENCE: 40170.60501
; CURRENT APPLICATION NUMBER: US/09/917,384
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 3687
; TYPE: DNA
; ORGANISM: Artificial Sequence

; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Segment of
; OTHER INFORMATION: Guxa
US-09-917-384-2

Alignment Scores:

Pred. No.: 6.37e-50 Length: 3687
Score: 823.00 Matches: 251
Percent Similarity: 47.07% Conservatives: 46
Best Local Similarity: 39.78% Mismatches: 173
Query Match: 16.03% Indels: 164
DB: 24
Gaps: 24

US-09-917-376-1 (1-957) x US-09-917-384-2 (1-3687)

Qy 426 TrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThr 445
Db 307 TGGATGACCGCATCGCTGCGATCAACGGCGTCAACGCG---GGACCC 351
Qy 446 GlyAlaThrLeuTyrAlaThrAsnAspLeuThrIleTyrAspSerGlyGlnIleHis 465
Db 352 GGCCTGACGACATATCTGGACGCGCGCTCTCCAG-----CAGCAGGGA 396
Qy 466 IleAlaProMetValLysGlyLeuGluThrAlaValAsnAspLeuIleSerProPro 485
Db 397 ACCACCCCTGAAGTC-----ATTGAGATTGTCTATCTACGATCTG---CCGGA 441
Qy 486 SerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaVal 505
Db 442 CGGACTGCGCGCGCTCGCTCAACGGCGAATGCGCGCT---ACGGCAGCAGTTTG 498
Qy 506 ThrAlaValProSerThrIlePheThrSerProValPheThrGlyThrSerValAsp 525
Db 499 CAGACCTAT---GAAACGCGAGTACATCGATTCGATTCGAGTATCTCAGCAATCCGA 555
Qy 526 TyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSerSerGln 545
Db 556 TACTCCAGCTCGCGATCGTGACGATCAT-----GAGCCGAGCTCGCTG 600
Qy 546 ProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySer 565
Db 601 CCAACCG-GGTAC-----CAATAT 620
Qy 566 GluProGlyGlyValThrThrGlyThrValAla-----AlaSerAlaAspGly 582
Db 621 GAGCATTCACGCGTGCACGCGGCTGCGCTATT-ACGAGCAAGGCATCGAGTACGCGC 679
Qy 583 SerArg-----PheValTrpAlaProGlyAspPro--- 592
Db 680 TCAGAAATTGACGCCCATTCGGAACGCTGATCATCTACATGACGCGCGCCACTCGGCT 739
Qy 593 -----GlyGlnProValVal-----TyrAlaValGlyPheGlyAsnSerTrp 606
Db 740 GGCTTGGTGGCCCAATAATGCCAGCGGATACGTACAGGAAGTCCAGAGGTCTCTCAACG 799
Qy 607 AlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsn----- 624
Db 800 CGAGCATCGGGGTCAACGGCATCGAGCTTCGTCACCAACACGCGGGAATTACACGCCCT 859
Qy 625 -----ProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyr 637
Db 860 TGAAGGAGCGCTTCATGACCGCCACCCAGAGGTGCGCGGACAGCGGTGGAGTCCGCGA 919
Qy 638 ArgSerThrAspGlyGlyValThrPheGlnPro----- 648
Db 920 ATTTCTACCAAGTGGAAATCTCGACATC-GACGAAGCCGACTACGCGGTGACTTGTACTCG 978
Qy 649 -----ValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaVal 666
Db 979 CGGCTCGTCCGCGCTGGCTTCCAGGAGC-----ATCGGATGCTCTATCGACACCTTA 1032
Qy 667 ProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSerThrAsn 686

Db 1033 -----CGCAAC 1038
Qy 687 GlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnVal----- 703
Db 1039 GGTGGGGTGGTCCGAACGAACCAACAGGCGCCGAGCAGCGGATGTCAACACCTTC 1098
Qy 704 -----Gly 704
Db 1099 GTCAACCACTCGAAGATTGACCTTCGCGCAGCAGCGCGCTGTGGTGCAACCAAGACGGT 1158
Qy 705 PheGlyValSerAlaProGlySerSerTrpProAlaValPhe----- 718
Db 1159 GCGGGCTCGCGCAGCGCGCGAGCAAGCCGACGACTTCGCAACGCGACCTCGAC 1218
Qy 719 -----ValValGlyThrIleGlyGlyValThrGlyAla----- 729
Db 1219 GCGTATGTCGATCAAGCGCGCGGTGAGTCGAGCGGCAAGCGGTGCGAGGATCCG 1278
Qy 730 -----TyrArgSerAspAsp---CysGlyThrThrTrpValLeuIleAsnAspAsp 745
Db 1279 ACAACTGGCAAGAGTCGAGCCCATGTCGAGCCCGAGGTACACGAG----- 1326
Qy 746 GlnHisGlnTyrGlyAsnTrpGlyGlnAla-----IleThrGlyAspHis 760
Db 1327 -----TCGTACGGGTACTGACCAACGCGTTTACCGAACTCCCGGATCGCGCGCAGTGG 1380
Qy 761 AlaAsnLeuArgArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIle 780
Db 1381 TTCCCGGCGAGTTTGCACGACTTGTGCGAAACGACGCGGTCGCGAGCTCGAC 1440
Qy 781 GlyGly-----AlaProSerGlySerProSerProSerValSer 793
Db 1441 AGTCGAGCGCGCGCTCCGCGCGAGTCCGTTCGCGCGAGTCCGAGCCCGAGT 1500
Qy 794 ProSerAlaSerProSerLeuSerProSerProSerProSerProSerProSer 813
Db 1501 CCGAGCCGAGCAGCTCGCCATCGCGCTCGCGCTCTCGAGCTCGAGCCCGTCTCGCTCG 1560
Qy 814 Pro-----SerProSerSerSerProSerProSerProSer 825
Db 1561 CCGAGCCGAGTCCGAGCCGAGTAGCTCGCGTCCGCGTCTCGAGCTCGAGCCCGTCT 1620
Qy 826 ProSerProSerProSerProSerProSerProSerProSerProSerProSer 845
Db 1621 CCGTCCCGAGCCGAGTCCGAGCCGAGTAGCTCGCGTCCGCGTCTCGAGCTCGAGC 1680
Qy 846 ProSerSerProSerProSerProSerProSerProSerProSerProSer 865
Db 1681 CCGTCTCCGTCGCGAGCCGAGTCCGAGCCGAGTAGCTCGCGTCCGCGTCTCGAGC 1740
Qy 866 SerSerProValSerGlyGlyValValValGlnTyrIleAsnAsnAspSerAlaProGly 885
Db 1741 TCGTCCCGGTGTCGGGTGGGTGAAGGTGAGTGTGAAATACCGGGTCTCGTGGGATTG 1800
Qy 886 AsnAsnGlnIleValProGlyLeuGlnValValValValValValValValValVal 905
Db 1801 GATAACCAATCAACCGGGTCTCCAGTGTGTAATACCGGGTCTCGTGGGATTG 1860
Qy 906 SerThrValThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyr 925
Db 1861 TCGAGGTGACGGTGGGTGCTGTTTCCCGGGATGGTGGTGGTGGTGGTGGTGGTGGT 1920
Qy 926 AsnCysAspTrpAlaAlaIleGlyCysGlyAsnIleArgAlaSerPheGlySerValAsn 945
Db 1921 AACTGTGACTGGCGCGGATGGGTGGGATATCCGCGCTCGTTCGGTGGTGAAC 1980
Qy 946 ProAlaThrProThrAlaAspThrTyrLeuGln 956
Db 1981 CCGGCGACCGCGGAGCACCTACCTGACAG 2013

RESULT 15

US-09-917-383-2

; Sequence 2, Application US/09917383

; Publication No. US20030104522A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOXYLICUS
; FILE REFERENCE: 40170.6US01
; CURRENT APPLICATION NUMBER: US/09/917,383
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 3687
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Segment of
; OTHER INFORMATION: Guxa
US-09-917-383-2

Alignment Scores:

Pred. No.:	6.37e-50	Length:	3687
Score:	823.00	Matches:	251
Percent Similarity:	47.07%	Conservative:	46
Best Local Similarity:	39.78%	Mismatches:	173
Query Match:	16.03%	Indels:	164
DB:	10	Gaps:	24

US-09-917-376-1 (1-957) x US-09-917-383-2 (1-3687)

Qy 426 TrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThr 445
Db 307 TGGATGAGCGCATCGCTCGCATCAACGCGCTCAACGCG-----GGACCC 351
Qy 446 GlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHis 465
Db 352 GCGTTGACGACATATCTGGACGCGCGCTCTCCGAG-----CAGCAGGGA 396
Qy 466 IleAlaProMetValLysGlyLeuGluThrAlaValAsnAspLeuIleSerProPro 485
Db 397 ACCACCTCGAGTC-----ATTGAGATTGTCATCTACGATCTG-----CGGGA 441
Qy 486 SerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspVal 505
Db 442 CGCGACTCGCGCGCTCGCTCCCAACGCGAACTGCCCGCT---ACGCGACGAGTTTG 498
Qy 506 ThrAlaValProSerThrIlePheThrSerProValPheThrGlyThrSerValAsp 525
Db 499 CAGACCTAT---GAAACGCGAGTACATCGATTCGATTCGAGGATCTCTGAGCAATCGGAAG 555
Qy 526 TyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSerSerGln 545
Db 556 TACTCCAGCTCGGATCGTGACCATCT---GAGCGGACTCGCTG 600
Qy 546 ProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySer 565
Db 601 CCAACGCG-GGTAC-----PheValTrpAlaProGlyAspPro---CAATAT 620
Qy 566 GluProGlyValThrThrGlyGlyThrValAla-----AlaSerAlaAspGly 582
Db 621 GAGCATTCAGCGGTGTCGAACGCGCGTCCGCTATT-ACGAGCAAGGATTCGAGTACGCGC 679
Qy 583 SerArg-----PheValTrpAlaProGlyAspPro--- 592
Db 680 TCAGAAATTCACGCCCATTCGAGCGGTACATCTACATGAGCGCGCCACTCGGGCT 739
Qy 593 -----GlyGlnProValVal-----TyrAlaValGlyPheGlyAsnSerTrp 606
Db 740 GGCTTGGTGGCCCAATAATGCCAGCGGATACGTACAGGAGTCCAGAAGGTCTCTCAACG 799

QY 607 AlaAlaSerGlnGlyValProAlaAlaAsnAlaGlnIleArgSerAspArgValAsn----- 624
 Db 800 CGAGCATCGGGTCACACGGCTTCACCGGCTTCACCAACACGGCGAATTCACACCGCGT 859
 QY 625 -----ProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyr 637
 Db 860 TGAAGAGCGGTCATGACCGCCACCGAGCGGTCCGCGGACACCGCGTGGAGTCGGCGA 919
 QY 638 ArgSerThrAspGlyValThrPheGlnPro----- 648
 Db 920 ATTCTACCAAGTGAATCCTGACATC-GACGAAGCGGACTACCGGGTTGACTTGTACTCG 978
 QY 649 -----ValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaVal 666
 Db 979 CGGCTCGTCCCGCTGCTTTCACAGCAGC-----ATCGGCATGCTTCATCGACACCTTA 1032
 QY 667 ProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSerThrAsn 686
 Db 1033 -----CGCAAC 1038
 QY 687 GlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnVal----- 703
 Db 1039 GGTGGGGTGGTCCGAACGACCAACAGGCCCGGACACCGCGCATGTCAACACCTTC 1098
 QY 704 -----Gly 704
 Db 1099 GTCAACAGTCGAAGATTGACCTTCGGCAGCAGCGCGCTGTGTGCAACAGAACGGT 1158
 QY 705 PheGlyLysSerAlaProGlySerSerTyrProAlaValPhe----- 718
 Db 1159 GCGGGCTCGGCGAGCGCGCGAGCGCAAGCCCGGACTTCCCGAACGCGCACCTCGAC 1218
 QY 719 -----ValValGlyThrIleGlyGlyValThrGlyAla----- 729
 Db 1219 GCGTATGTCGGATCAAGCCCGCGGTGAGTCGGACGCGCAAGCGCTGGAGCGATCCG 1278
 QY 730 -----TyrArgSerAspAsp-----CysGlyThrThrTrpValLeuIleAsnAspAsp 745
 Db 1279 ACAACTGGCAAGAGTCGAGCCCGCATGTGCGACCGCGAGCTACACGAG----- 1326
 QY 746 GlnHisGlnTyrGlyAsnTrpGlyGlnAla-----IleThrGlyAspHis 760
 Db 1327 -----TCGTACGGGGTACTGACCAACCGCGTTTACCGAACTCCCGCATCGCGCGCAGTGG 1380
 QY 761 AlaAsnLeuArgArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIle 780
 Db 1381 TTCCCGCGCGAGTTTGACAGCTGTTCGCGAAGCAGCGCGAGCGGTGCGAGCTCGAC 1440
 QY 781 GlyGly-----AlaProSerGlySerProSerProSerValSer 793
 Db 1441 AGCTCGAGCCCGCGCTCCGCGCGAGTCCGTTCGCGCTTCGCGAGTCCGAGCCCGAGT 1500
 QY 794 ProSerAlaSerProSerLeuSerProSerProSerProSerSerSerProSerProSer 813
 Db 1501 CCAGGCCCGAGCAGCTCGCCATCGCGCTCGCGTCTCCGAGCTCGAGCCCGTCTCCGTCG 1560
 QY 814 Pro-----SerProSerSerProSerProSerProSerProSer 825
 Db 1561 CCAGGCCCGAGTCCGAGCCCGAGTAGCTCGCCGTTCGCGCTTCGAGCTCGAGCCCGTCT 1620
 QY 826 ProSerProSerProSerProSerArgSerProSerProSerProSerAlaSerProSer 845
 Db 1621 CCGTCGCGGAGCCCGAGTCCGAGCCCGAGTAGCTCGCGCTCGCGTCTCCGAGCTCGAGC 1680
 QY 846 ProSerSerProSerProSerProSerSerSerProSerProSerProSerProSer 865
 Db 1681 CCGTCTCCGTCGCGAGCCCGAGTCCGAGCCCGAGTAGCTCGCGTCTCGCGTCTCCGAGC 1740
 QY 866 SerSerProValSerGlyValLysValGlnTyrIleAsnAsnAspSerAlaProGly 885
 Db 1741 TCGTCGCGGTGTGGGTGGGCTGAAGGTGAGTACGAAGAACATGATTCGCGCGCGGGT 1800
 QY 886 AspAsnGlnIleLysProGlyLeuGlnValValAsnThrGlySerSerValAspLeu 905

Db 1801 GATAACAGATCAAAACCGGGTCTCCAGTTGGTGAATACCGGGTCTCGTCGGTGGATTG 1860
 QY 906 SerThrValThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyr 925
 Db 1861 TCGACGGTGACGGTGGGTACTGGTTCAACCGGGATGGTGGTCTCGACACTGGGTGAC 1920
 QY 926 AsnCysAspTrpAlaAlaIleGlyCysGlyAsnIleArgAlaSerPheGlySerValAsn 945
 Db 1921 AACTGTGACTGGCGCGGATGGGGTGTGGGAATATCCGCGCTCGTTGGCTCGGTGAAC 1980
 QY 946 ProAlaThrProThrAlaAspThrTyrLeuGln 956
 Db 1981 CCGCGCAGCGCGACCGCGGACACCTACCTGCAG 2013

Search completed: May 15, 2004, 03:52:53
 Job time : 13740 secs